

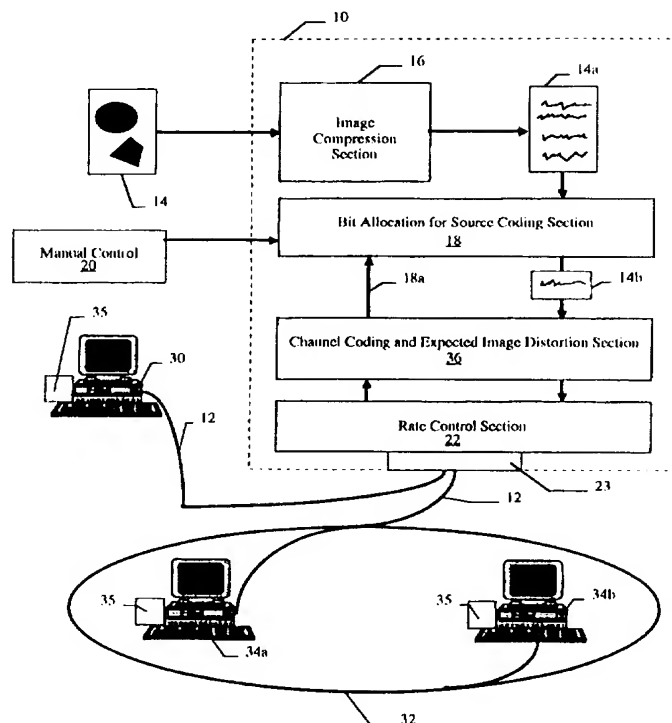


## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>H04N 1/41</b>		A1	(11) International Publication Number: <b>WO 97/21302</b>
			(43) International Publication Date: 12 June 1997 (12.06.97)
(21) International Application Number: PCT/US96/19388		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 5 December 1996 (05.12.96)			
(30) Priority Data:			
60/008,294	8 December 1995 (08.12.95)	US	
60/023,569	7 August 1996 (07.08.96)	US	
60/024,804	29 August 1996 (29.08.96)	US	
(71) Applicant (for all designated States except US): TRUSTEES OF DARTMOUTH COLLEGE [US/US]; 11 Rope Ferry Road, Hanover, NH 03755 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): DANSKIN, John, M. [US/US]; 12 Fletcher Circle, Hanover, NH 03755 (US). DAVIS, Geoffrey [US/US]; Apartment 2W, 26 E. Wheelock, Hanover, NH 03755 (US).			
(74) Agents: VOCK, Curtis, A. et al.; Lappin & Kusmer L.L.P., Two Hundred State Street, Boston, MA 02109 (US).			

**Published***With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.***(54) Title:** FAST LOSSY INTERNET IMAGE TRANSMISSION APPARATUS AND METHODS**(57) Abstract**

Fast lossy Internet image transmission ("FLIIT") systems and methods are provided for transmitting images (14), such as world wide web graphics, over the Internet. Forward error correction, added to an image during compression, enables the subsequent reconstruction of fragments lost during transmission by purposefully concentrating image bits within the portions of the image that have the greatest overall visual impact. Image fragments that are lost during transmission have little noticeable effect, and no time is spent on retransmitting lost fragments, such as in TCP/IP. FLIIT eliminates retransmission delays by strategically shielding important parts of subband coded images through forward error correction. Each subband is decomposed into a series of bitplanes ordered from the most significant to the least significant. An optimization procedure determines the subset of bitplanes to transmit as well as the number of bits to spend on forward error correction for each bitplane, recognizing that different bits in compressed images such as JPEG have different contributions to image fidelity. FLIIT also assesses current network conditions and adjusts transmission rates so as to accommodate network traffic: keeping the total transmission bits constant, more bits are adjusted to data during low network congestion, while more bits are adjusted to redundancy during high network congestion. A decoding section within a receiver unit, e.g., personal computer, decodes the transmitted image upon arrival across the Internet, providing a 2-4 factor improvement in speed over existing image transfers with the same quality.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

- 1 -

# Fast Lossy Internet Image Transmission Apparatus and Methods

## Related Applications

This is a continuing application of Provisional Application No. 60/008,294, filed on December 8, 1995, and entitled "Fast Lossy Internet Image Transmission Apparatus and Methods," of Provisional Application No. 60/023,569, entitled "Fast Lossy Internet Image Transmission Apparatus and Methods" and filed on August 6, 1996, and of Provisional Application No. 60/024804, entitled "Fast Lossy Internet Image Transmission Apparatus and Methods" and filed on August 29, 1996, each of which is hereby incorporated by reference.

## Background

World Wide Web requests are the single largest consumer of Internet bandwidth, comprising roughly 25% of all bytes sent. *See*, Georgia Tech Graphics, Visualization, & Usability Center, "Third degree polynomial curve fitting for bytes transferred per month by service," NSFNET Backbone Statistics Page, August 1995, <http://www.cc.gatech.edu/gvu/stats/NSF/-merit.html>. Images, most of which are examined for only a few seconds, undoubtedly constitute the bulk of the ten terabytes of current monthly Web requests. For such interactive applications as web browsers, the responsiveness gained from rapid image transmission is more important than perfect image fidelity, since many images are already distorted by lossy compression, and since relatively few images are closely examined.

The usual method for transmitting images over the Internet is to first compress the images using a lossy scheme such as JPEG, and then to transmit

-2 -

1 the compressed images across the intrinsically lossy Internet using the lossless  
2 TCP/IP protocol. JPEG and related lossy schemes are very sensitive to bit errors  
3 and hence require lossless transmission. The price paid for lossless  
4 transmission over a lossy medium, however, is excessively lengthy  
5 transmission times due to retransmissions of lost packets.

6  
7 Specifically, TCP/IP retransmits missing pieces until the image is  
8 complete, resulting in inefficiencies and considerable transmission delays.  
9 This is particularly true with the growing popularity of the Internet, which has  
10 led to increased network congestion and "traffic jams" that cause fragments of  
11 images to be lost in transit. Because lossless TCP/IP depends upon  
12 retransmission to correct network losses, the transmission time for even  
13 relatively short messages can be substantial, particularly during times of heavy  
14 network traffic.

15  
16 Lossless transmission schemes are even more problematic for Internet  
17 video broadcasting. Retransmission is impractical with such broadcasting  
18 because the receivers will not, in general, experience the same losses.  
19 Accordingly, a broadcaster attempting to respond to all of the different losses  
20 can be overwhelmed with requests to retransmit lost packets.

21  
22 A number of strategies have been explored for incorporating  
23 redundancy into network packets. In Turner et al., Image transfer: an end-to-  
24 end design, SigComm 92, 258-268, for example, a scheme is presented in which  
25 errors are corrected by making use of naturally occurring redundancy within  
26 images. Image pixels are reordered for transmission in such a way that packet  
27 losses cause the loss of isolated pixels rather than of large contiguous blocks of  
28 pixels. Missing pixels are reconstructed by applying a filter to neighboring  
29 pixels that survive transmission, thereby hiding a limited number of missing

1 packets when there is high correlation between neighboring pixels. However,  
2 when such a correlation does not exist, the technique does not readily mask  
3 missing packets.

4  
5 In a related scheme, a network video transmission scheme proposed by  
6 Karlsson et al., Subband coding of video for packet networks, Optical  
7 Engineering, 27(7), 574-586 (1988), also makes use of naturally occurring image  
8 redundancy for error correction. As above, using intrinsic image redundancy  
9 to correct losses remains problematic, since the number of losses that can be  
10 sustained is highly image dependent. Furthermore, when efficient  
11 compression schemes are used, very little usable redundancy remains for error  
12 correction. That is, common image compression techniques typically operate to  
13 remove redundant image pixels so that reconstruction of adjacent pixels is  
14 ineffective.

15  
16 In other techniques, the control of transmission errors is obtained by  
17 adding redundancy bits to the bitstream rather than by relying solely on  
18 naturally occurring redundancy. In Biersack, Performance evaluation of  
19 forward error correction in ATM networks, Proceedings of the SIGCOMM 92  
20 Symposium, Baltimore, 248-257 (1992), for example, a technique is presented  
21 that evaluates the effect of redundancy addition at a fixed rate to video  
22 transmissions over ATM networks. This fixed rate addition of redundancy,  
23 however, is inherently inefficient and can obtain mixed results. By way of  
24 example, testing of this technique has shown that in heterogeneous traffic  
25 scenarios, the loss rates were reduced by several orders of magnitude; but for  
26 more homogeneous traffic scenarios, the performance was unchanged or  
27 worsened. Further, in homogeneous traffic scenarios, the increase in the  
28 network load from the transmission of redundancy bits can cause an increase  
29 in the loss rate not compensated for by the error correction.

1  
2 Another prior art method of adding redundancy is through joint source  
3 and channel coding. C. Shannon, in A Mathematical Theory of  
4 Communication, Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656 (  
5 1948), describes a source-channel separation theorem which states that separate  
6 source and channel coding procedures can be made to be just as effective as a  
7 joint procedure. Nevertheless, the results of joint source and channel coding  
8 are asymptotic and require infinite length messages.  
9

10 A Lagrange multiplier-based joint source-channel coding scheme for  
11 continuous bitstreams has also been developed in the prior art. See, e.g., N.  
12 Tanabe et al., Subband image coding using entropy-coded quantization over  
13 noisy channels, IEEE Journal on Selected Areas in Communications, 10:5, 926-  
14 943 (1992). In this scheme, however, error calculations for continuous streams  
15 are extremely complex, and the algorithms presented rely on computationally  
16 expensive simulations during bit allocation.  
17

18 A related source-channel coding scheme for networks, entitled "Priority  
19 Encoding Transmission" (hereinafter "PET"), has also been developed in the  
20 prior art. See A. Albanese et al., Priority encoding transmission, Proc. 35th  
21 Annual Symposium on Foundations of Computer Sciences, Santa Fe, NM, pp.  
22 604-612 (1994); C. Leicher, Hierarchical encoding of MPEG sequences using  
23 priority encoding transmission (PET), TR-94-058, ICSI, Berkeley, CA (1994). The  
24 implementation of PET for MPEG allows the user to set different levels of error  
25 protection for different portions of the MPEG stream, but provides little or no  
26 methodology for allocating these levels.  
27

28 The layered transmission schemes in M. Garrett, Joint source/channel  
29 coding of statistically multiplexed real-time services on packet networks, IEEE

-5 -

1 Transactions on Networking, 1:1, 71-80 (1993), and E. Posnak et al., Techniques  
2 for resilient transmission of JPEG video streams, also make use of prior art  
3 joint source-channel coding methods. These layered schemes require  
4 networks that treat packets differently according to their priorities. Visually  
5 important data is thus sent with a high priority, i.e., with a smaller loss rate,  
6 and less important data is sent with a low priority so as to be discarded first by  
7 switches during congestion. Accordingly, these schemes require networks  
8 capable of providing prioritized handling of packets, a capability that is not  
9 always available on the Internet.

10  
11 Another prior art lossless Internet flow control technique is described in  
12 L. Brakmo, TCP Vegas: New techniques for congestion detection and  
13 avoidance, Proceedings of the SIGCOMM '94 Symposium, (1994). This TCP  
14 Vegas technique achieves rate control by managing the number of packets  
15 stored in the network, rather than by forcing losses as TCP Reno does.  
16 However, this technique is also problematic in that it has a relatively slow  
17 start-up time.

18  
19 It is desirable to speed up Internet image transmissions so that online  
20 users can view, download and operate on images more quickly. This is true  
21 regardless of the bandwidth of the online connection, e.g., an Internet  
22 connection including a personal computer, modem and phone line operating  
23 at 28.8kbps (kilobytes-per-second), a digital telephone line, and/or the ethernet.  
24 It is especially desirable to increase the speed of Internet image transmission  
25 without significant loss in image quality.

26  
27 It is, accordingly, an object of the invention to provide apparatus and  
28 methods for increasing the speed of Internet image transmissions.

1  
2           Still another object of the invention is to provide systems and methods  
3 for adjusting the transmission speed of Internet images with selectable image  
4 quality.

5  
6           Yet another object of the invention is to provide a fast, lossy Internet  
7 image transmission methodology which reduces the problems associated with  
8 prior art Internet image transmission methods.

9  
10          Another object of the invention is to provide an error correction system  
11 which speeds up Internet image transmissions in a manner compatible with  
12 existing networks and without significant loss of image quality.

13  
14          These and other objects will become apparent in the description which  
15 follows.

16



## 1 Summary of the Invention

2  
3 The invention provides an efficient method for transmitting images,  
4 such as world wide web graphics, over the Internet. In a preferred aspect, the  
5 invention makes use of forward error correction ("FEC"), which allows the  
6 recipient of an image on the Internet to reconstruct fragments lost during  
7 transmission. Preferably, the FEC methodology of the invention is added to an  
8 image during compression, and purposefully concentrates image bits within  
9 the portions of the image that have the greatest overall visual impact.  
10 Accordingly, image fragments that are lost during transmission have little  
11 noticeable effect, and no time is spent on retransmitting lost fragments, such as  
12 in TCP/IP.

13  
14 More particularly, in one aspect, the invention provides a fast lossy  
15 Internet image transmission (hereinafter "FLIIT") methodology that  
16 eliminates retransmission delays by strategically shielding important parts of  
17 subband coded images through FEC. Each subband is decomposed into a series  
18 of bitplanes ordered from the most significant to the least significant. An  
19 optimization procedure, described in more detail below, determines the subset  
20 of bitplanes to transmit as well as the number of bits to spend on FEC for each  
21 bitplane. Bits are allocated in order to maximize the expected quality of the  
22 received image subject to an overall bit budget. The FLIIT methodology  
23 recognizes that different bits in compressed images such as JPEG have different  
24 contributions to image fidelity. For example, flipping high order bits in the DC  
25 channel of a JPEG-compressed image results in a large discernible difference in  
26 the decompressed image, whereas flipping low order bits in a high frequency  
27 channel has little visual effect. Typically, applying equal amounts of  
28 redundancy to protect bits in these two categories is not efficient.

1  
2       The FLIIT methodology preferably utilizes a first order Markov model of  
3 the bursty Internet packet loss structure. The use of the Markov model enables  
4 the determination of the effects of network burst errors within parity groups.  
5

6       In another aspect, the invention incorporates error correction into a  
7 standard wavelet-based subband coder. Specifically, the FLIIT methodology of  
8 the invention allocates bits between the tasks of encoding image subbands and  
9 protecting coded data with FEC. Bits devoted to subband coding correspond to  
10 the image to be transmitted, and bits devoted to FEC increase the likelihood of  
11 that image arriving intact. This allocation reduces the distortions in the  
12 received image, both from compression and network losses, and subject to a  
13 constraint on the total bytes transmitted. Accordingly, the FEC bits are  
14 concentrated in subbands where losses would be visually catastrophic, while  
15 less important subbands receive less protection.  
16

17       In still other aspects, the invention addresses network issues such as  
18 rate, congestion control, and startup. FLIIT methodology allocates a fixed  
19 number of bits between redundancy and data depending on the expected loss  
20 rate. When the loss rate is high, bits are shifted from data to redundancy, but  
21 the total number of transmitted bits remains constant. In the prior art, on the  
22 other hand, TCP retransmits more and more packets during heavy congestion  
23 because packet loss rates are high, presenting a positive feedback that worsens  
24 congestion. In accord with the invention, FLIIT methodology reduces this  
25 positive feedback by sending packets with a fixed total number of bits exactly  
26 once, trading quantizer resolution for FEC as a function of current network  
27 conditions.  
28

-9 -

1 Unlike the prior art, one aspect of a system constructed according to the  
2 invention includes a server that remembers the last sending rate for each  
3 recent connection, eliminating slow startups for repeat connections. Prior art  
4 TCP, on the other hand, has a slow startup procedure can take seconds to ramp  
5 up to full speed, which can be quite ineffective, particularly with respect  
6 compressed images like JPEGs. The invention achieves flow control by  
7 managing the number of packets stored in the network, and usually avoids the  
8 slow startup associated with TCP by remembering transmission rates for recent  
9 connections so that multiple short connections to a single client will only have  
10 to pay for one startup.

11  
12 Another aspect of the invention includes a data determination section  
13 that evaluates and decides when to stop waiting for data packets that may have  
14 been lost or delayed. The data determination section monitors and assesses the  
15 waiting period for packets, and balances that wait period between an  
16 insufficient time that risks losing image data, and an excessive time that  
17 results in reduced system responsiveness.

18  
19 In another aspect, the invention includes a burst-loss control module  
20 within a Bit Allocation for Source Coding Section which interleaves packets in  
21 order to decorrelate burst losses, thereby advantageously utilizing the structure  
22 of burst losses to achieve improved transmission fidelity.

23  
24 Further, in another aspect, the invention can include a subband coder  
25 module within a Channel Coding and Expected Image Distortion Section using  
26 nested quantization to reduce FEC requirements.

27  
28 In still another aspect, the invention includes a flow control section  
29 which operates to choose an optimal time to stop waiting for lost packets.

-10 -

1 Typically, the ideal stopping point is the expected time of arrival of the last  
2 packet plus the standard deviation of the interpacket arrival time.

3  
4 In still another aspect, the invention provides a system which  
5 efficiently transmits image data via some redundant transmission such as FEC,  
6 which makes serious transmission errors unlikely. The system applies this  
7 redundancy selectively, however, since the addition of redundancy increases  
8 the amount of information that must be transmitted. Indeed, experiments in  
9 FEC redundancy as applied uniformly to ATM video packets have shown a  
10 decrease in performance in some cases since the increased network load due to  
11 the redundancy can lead to an increase in the packet loss rate. See Biersack,  
12 Performance evaluation of forward error correction in ATM networks,  
13 Proceedings of the SIGCOMM 92 Symposium, Baltimore, 248-257 (1992).  
14 Accordingly, the invention provides certain operational controls, described in  
15 more detail below, which function to optimize FEC redundancy in view of  
16 network congestion, desired image quality, and/or transmission speed.

17  
18 The invention thus provides several important advantages over the  
19 prior art. First, the invention speeds up Internet image transmissions by a  
20 factor of 2 to 4 over TCP while maintaining images of similar overall quality.  
21 The invention is also suitable for fast transmission of video over the Internet,  
22 and, more importantly, seamlessly coexists with existing TCP connections.

23  
24 The invention further provides a methodology for obtaining an  
25 optimized partitioning of bits between source coding and channel coding for a  
26 given set of (1) image subband quantizers, (2) FEC protection levels, and (3)  
27 packet loss model. In one aspect, for example, a rate control section is provided  
28 to accommodate other lossy Internet media protocols, such as real time voice  
29 transmission.

1  
2 The following articles and book chapters provide useful background to  
3 the invention and are, accordingly, incorporated herein by reference: A.  
4 Albanese et al., Priority encoding transmission, Proc. 35th Annual Symposium  
5 on Foundations of Computer Sciences, Santa Fe, NM, pp. 604-612 (1994); T.C.  
6 Bell et al., Text Compression, Prentice Hall, Englewood Cliffs, NJ (1990); E.W.  
7 Biersack, Performance evaluation of forward error correction in ATM  
8 networks, Proceedings of the SIGCOMM 92 Symposium, Baltimore, 248-257,  
9 (1992); L. S. Brakmo et al., TCP Vegas: New techniques for congestion detection  
10 and avoidance, Proceedings of the SIGCOMM '94 Symposium (1994); T.M.  
11 Cover et al., Elements of Information Theory, John Wiley & Sons, Inc., New  
12 York (1991); J. M. Danskin, G. Davis, and X. Song, "Fast Lossy Internet Image  
13 Transmission," ACM MultiMedia 95, pp. 321-332 (1995); B. Fox, Discrete  
14 optimization via marginal analysis, Management Science 7 13:3, pp. 210-216  
15 (1966); G. Karlsson et al., Subband coding of video for packet networks, Optical  
16 Engineering, 27(7), 574-586 (1988); C. Leicher, Hierarchical encoding of MPEG  
17 sequences using priority encoding transmission (PET), TR-94-058, ICSI,  
18 Berkeley, CA ( 1994); W. E. Leland et al., On the self-similar nature of ethernet  
19 traffic, Proc. SIGCOMM, 183-193, San Francisco (1993); A. S. Lewis et al., Image  
20 compression using the 2-D wavelet transform, IEEE Transactions on Image  
21 Processing, Vol. 1, No. 2, pp. 244-250 (1992); C. E. Shannon, A Mathematical  
22 Theory of Communication, Bell System Technical Journal, Vol. 27, pp. 379-423,  
23 623-656 ( 1948); J. Shapiro, Embedded Image Coding Using Zerotrees of Wavelet  
24 Coefficients, IEEE Transactions on Signal Processing, Vol. 41, No. 12, pp. 3445-  
25 3462; Y. Shoham and A. Gersho, Efficient bit allocation for an arbitrary set of  
26 quantizers, IEEE Trans. Acoustics, Speech, and Sig. Proc., 36:9, 1445-14537  
27 (1988); N. Tanabe, Subband image coding using entropy-coded quantization  
28 over noisy channels, IEEE Journal on Selected Areas in Communications, 10:5,  
29 926-943, (1992); A. Tanenbaum, Computer Networks, Prentice-Hall, Englewood

1 Cliffs, N. J. (1981); Turner, Charles J., and Larry L. Peterson, "Image transfer:  
2 and end-to-end design," SigComm 92, 258-268; D. Taubman, Multirate 3-D  
3 subband coding of video, IEEE Trans. Image Proc., 3(5) (1994); J.D. Villasenor et  
4 al., Wavelet filter evaluatio for image compression, IEEE Trans. Image  
5 Processing (1995); C.L. Williamson et al., Loss-load curves: Support for rate-  
6 based congestion control in high-speed datagram networks, Proceedings of  
7 SIGCOMM 91, pp. 17-287 (1991); I. Witten et al., Arithmetic coding for data  
8 compression, Communications of the ACM, 30:6, 520-540 (1987).

9  
10 The invention is next described further in connection with preferred  
11 embodiments, and it will become apparent that various additions, subtractions,  
12 and modifications can be made by those skilled in the art without departing  
13 from the scope of the invention.

#### 14 15 Brief Description of the Drawings

16  
17 A more complete understanding of the invention may be obtained by  
18 reference to the drawings, in which:

19  
20 Figure 1 shows a schematic layout of a system constructed according to  
21 the invention;

22  
23 Figure 1A shows a schematic layout of another system constructed  
24 according to the invention, including software modules to enable FLIIT  
25 methodology;

26  
27 Figure 1B illustrates blocks of data sorted by typical redundancy, in  
28 accord with the invention;

29

-13 -

1           Figure 1C illustrates message transmission through typical routers on  
2 the Internet;

3  
4           Figure 1D illustrates a graphical distribution of bitplanes and parities for  
5 a representative image in accord with the invention;

6  
7           Figure 2 illustrates a flowchart for encoding images in accord with the  
8 invention;

9  
10          Figure 2A illustrates a flowchart for decoding images in accord with the  
11 invention;

12  
13          Figure 2B illustrates an alternative flowchart for encoding images in  
14 accord with the invention;

15  
16          Figure 2C graphically shows Internet packet drop rate for packets sent  
17 between Dartmouth College and Stanford University as a function of time of  
18 day;

19  
20          Figure 2D graphically shows observed and fitted cumulative density  
21 functions for packet delays modeled according to the invention;

22  
23          Figure 3 graphically illustrates the expected and measured PSNR  
24 performances of FLIIT methodology, according to the invention, and three  
25 fixed parity schemes;

26  
27          Figures 4A-4C shows experimental results of Lena images transmitted  
28 over a transcontinental Internet connection utilizing flat parity schemes with  
29 differing image quality reconstruction percentiles;

-14 -

Figures 5A-5C shows experimental results of Lena images transmitted over a transcontinental Internet connection utilizing FLIIT methodology, with differing image quality reconstruction percentiles, according to the invention;

Figure 5D illustrates the Lena image of Figures 4 and 5 transmitted via TCP/IP;

Figure 6 schematically illustrates one test configuration used to test the system of the invention;

Figure 7 graphically shows the time advantages of transmitting images via FLIIT as compared to TCP for selected image qualities; and

Figures 8A and 8B graphically show the time impact of FLIIT vs. TCP protocols for various network configurations.

#### Detailed Description of Illustrated Embodiments

Figure 1 illustrates a system 10 constructed according to the invention for transmitting images through the Internet 12. An uncompressed electronic image 14 is first reduced in size by an Image Compression Section 16 so as to produce, for example, lossy JPEG representations 14a of the image 14. The Bit Allocation for Source Coding Section 18 thereafter partitions and transforms the image 14a into a set of subbands ranging from high frequency, fine scales to low frequency, coarse scales so as to minimize image distortions relative to a total allowed number of transmission bits. These transmission bits form the electronic image file 14b with finely quantized coefficients that contribute heavily to image fidelity, e.g., low frequency image components, and coarsely



-15 -

1 quantizing coefficients that contribute little to image fidelity, e.g., high  
2 frequency edges.

3  
4 File 14b is thus suitable for transmission through the Internet 12 and to a  
5 client receiving unit 30, e.g., a personal computer, and/or through the Internet  
6 12 and into a network 32 that includes a plurality of client receiving units 34a,  
7 34b. Each of the units 30, 34a, 34b has a decoding subsection 35 housed within  
8 associated memory, e.g., firmware or application-specific software within  
9 random access memory ("RAM"). As described in more detail below, the  
10 decoding subsection 35 operates to "reverse" the encoding process provided by  
11 sections 16, 18, except that no bit allocation decisions are made and certain  
12 image packets are unavailable due to transmission losses along the Internet  
13 and network 12, 32, respectively.

14  
15 System 10 preferably includes a Channel Coding and Expected Image  
16 Distortion Section 36, described below, which dynamically allocates bits  
17 between source and channel codes depending upon conditions within the  
18 network 32. System 10 connects to the Internet 12 through any of the standard  
19 interfaces, e.g., an ethernet connection 23.

20  
21 System 10 can be implemented in several ways. Generally, however,  
22 system 10 includes a central processing unit ("CPU") and one or more  
23 connected memories, such as shown in Figure 1A. In Figure 1A, system 10' is a  
24 computer or server that includes an image compression section 16', a Bit  
25 Allocation for Source Coding Section 18', a Channel Coding and Expected  
26 Distortion Section 36', and a Rate Control Section 22, each of which represents  
27 a software module in active memory 21' within the computer 10'. System 10'  
28 connects to the Internet 12' through any one of the known prior art  
29 connections, e.g., an ethernet connection 23', and transmits images and

-16 -

1 receives packet information from any of the connected users 30' to adjust  
2 image transmission characteristics, as described herein. The CPU 27 controls  
3 the system 10', including the input and output of image files into internal  
4 memory 21'.

#### 6 Image Compression Section

8 Image compression such as performed by the image compression section  
9 16, Figure 1, or section 16', Figure 1A, can occur by one of several methods. By  
10 way of example, sections 16, 16' can utilize a wavelet transform coding scheme  
11 to compress images for transmission along the Internet 12. Although those  
12 skilled in the art will appreciate that other compression schemes can be used,  
13 the wavelet-based coder is chosen and described herein because of its simplicity  
14 and excellent performance at low bit rates. Experimental results yield, without  
15 error correction overhead, peak signal-to-noise ratios (PSNR's) to within less  
16 than one dB of an embedded zerotree wavelet coder.

18 With further reference to Figure 1, a discrete wavelet transform is  
19 performed by the image compression section 16 on the image 14 by quantizing  
20 the coefficients using uniform quantizers, and by coding the resulting  
21 coefficients for entropy using an arithmetic coder. The resolution of the  
22 quantizers is determined by a Lagrange multiplier procedure or other  
23 optimization procedure describe in more detail below. One suitable transform  
24 is a 9/7-tap biorthogonal filter set used in experiments and as described in J.D.  
25 Villasenor et al., IEEE Trans. Image Processing (1995).

### Bit Allocation for Source Coding Section

The discrete wavelet transform performed by the Image Compression Section 16 results in a compressed image 14a. The Bit Allocation for Source Coding Section 18 thereafter partitions the image 14a into a set of subbands ranging from fine scales, i.e., high frequency, to coarse scales, i.e., low frequency. In natural images, the bulk of the visually important information is concentrated in the coarse-scale subbands, with the fine scale subbands contributing primarily to sharp edge effects. In accord with the invention, the Bit Allocation for Source Coding Section 18 transforms the image 14a into image representation 14b by finely quantizing coefficients that contribute heavily to image fidelity and coarsely quantizing others. Determining the quantization resolution of each subband is a key feature of the Bit Allocation for Source Coding Section 18. More particularly, section 18 performs a tradeoff between quantization error and total storage cost, and allocates quantizer resolutions to obtain minimal distortion for a given bit expenditure. The total bit expenditure can be set by two principal ways: through manual control 20, e.g., a computer and keyboard connected for communication with the system 10, or through feedback determinations of the Rate Control Section 22, each of which is described in more detail below.

The Bit Allocation for Source Coding Section 18 first selects one of a family of quantizers  $Q_0 \dots Q_k$  for each image subband. The quantizers are arranged from coarsest ( $Q_0$ ) to finest ( $Q_k$ ) and have bin widths scaled according to the range  $R_j$  of coefficients in each subband. By way of example, certain of the experiments described below employ quantizers  $Q_k$  with  $\{2^k - 1\}_{0 \leq k \leq 10}$  uniformly spaced bins. Quantizer bins are distributed symmetrically about 0, since wavelet coefficients are known *a priori* to be symmetrically distributed about the origin, and the bins for quantizer  $Q_k$  when quantizing subband  $j$

-18 -

1 have width  $2R_j / (2^k - 1)$ , where  $R_j$  is the maximum magnitude of a coefficient  
2 in subband  $j$ . Quantized values are preferably decoded to the center of each  
3 quantizer bin.

4  
5 In an alternative quantization, section 18 can utilize a family of  
6 quantizers such as described in D. Taubman et al., Multirate 3-D subband  
7 coding of video, IEEE Trans. Image Proc., 3(5) (1994), whereby a family of  
8 nested  $Q_k$  quantizers,  $2^k - 1$  bins, are used: one bin of width  $(2^{k+1}) R_j$  is centered  
9 at the origin; and the other  $2^k - 2$  bins are spaced uniformly and symmetrically  
10 around the center bin, each with width  $2^k R_j$ . This family of quantizers has the  
11 important property that quantizer bins are nested, i.e. each bin of  $Q_k$  can be  
12 decomposed into either two or three bins in  $Q_{k+1}$ . The output of the quantizer  
13  $Q_k$  can be expressed as a string of refinements  $(r_0, r_1, \dots, r_k)$ , where each of the  
14  $r_i$ 's is a 0, 1, or 2. The sets of refinements are essentially the bitplanes of the  
15 coefficients ordered from the most significant bit to the least significant bit.  
16 This family of nested quantizers permits fine control of the distribution of  
17 redundancy so as to vary the protection at the bitplane rather than the  
18 coefficient level.

19  
20 The Bit Allocation for Source Coding Section 18 also determines image  
21 distortion during the allocation of bits to the subbands. By way of example, a  
22 mean squared error function can be used to assess distortion. This choice also  
23 permits comparison with other algorithms; however, the mean squared error  
24 function functions equally well with perceptually weighted metrics such  
25 known to those in the art. See, e.g., S. Lewis et al., Image compression using the  
26 2-D wavelet transform, IEEE Transactions on Image Processing, Vol. 1, No. 2,  
27 pp. 244-250 (1992).

28

1        In particular, for the mean squared error function, let  $D_j(k)$  be the total  
 2 squared error incurred in quantizing the wavelet coefficients in subband  $j$  with  
 3 quantizer  $Q_k$ , and let  $C_j(k)$  be the cost in bits of representing the corresponding  
 4 entropy-coded quantized values. For an image decomposed into  $n$  subbands,  
 5 the Bit Allocation for Source Coding Section 18 computes to identify a vector  $\mathbf{q}$   
 6  $= (q_1, q_2, \dots, q_n)$  of quantizer indices so that the total distortion  $D_{\text{total}}(\mathbf{q}) =$   
 7  $\sum_{j=1}^n D_j(q_j)$  is minimized subject to the constraint that the total cost in bits,  
 8  $C_{\text{total}}(\mathbf{q}) = \sum_{j=1}^n C_j(q_j)$  is less than or equal to some given bit budget  $C_{\text{max}}$ . Section  
 9 18 thus seeks a minimization over  $\mathbf{q} \in Q$  where  $Q$  is a given set of valid  
 10 vectors of quantizer indices.

11  
 12        Marginal analysis, as known to those skilled in the art, see, e.g., B. Fox,  
 13 Discrete optimization via marginal analysis, Management Science 7 13:3, pp.  
 14 210-216 (1966), provides one algorithm suitable for solving this minimization  
 15 problem. In particular, the Bit Allocation for Source Coding Section 18  
 16 initializes the vector of quantizer resolutions  $\mathbf{q}$  to the coarsest configuration,  $(0,$   
 17  $0, \dots, 0)$ , and sets the number of remaining bits to allocate to  $C_{\text{max}}$ . Allocation  
 18 then proceeds iteratively as follows: for each subband, the cost and distortion  
 19 changes resulting from refining the subband's quantizer by one increment is  
 20 computed. All the subbands for which quantizer refinement is possible are  
 21 considered, providing that the cost of refinement does not exceed the total  
 22 remaining bits to allocate. If there are no such subbands, the Bit Allocation for  
 23 Source Coding Section 18 terminates the algorithm. Otherwise, it finds  
 24 subband  $j$  for which quantizer refinement yields the largest reduction in  
 25 distortion per bit, increments the corresponding  $q_j$ , and subtracts the cost of the  
 26 refinement from the total remaining bits.

-20 -

1 Marginal analysis in accord with the invention thus yields an optimal  
2 bit allocation when cost and distortion functions are convex. Marginal analysis  
3 is also very fast relative to the cost of the transform, requiring at most  $nK$   
4 iterations to converge, where  $n$  is the number of subbands and  $K$  is the number  
5 of quantizers in the family.

6  
7 Those skilled in the art should appreciate that other bit allocation  
8 techniques can be used in accord with the invention. For example, the  
9 minimization problem solved by Section 18 over  $\mathbf{q} \in Q$  can be solved by  
10 Lagrangian techniques as opposed to the marginal analysis described above. In  
11 Shoham et al., Efficient bit allocation for an arbitrary set of quantizers, *IEEE*  
12 *Trans. Acoustics, Speech, and Sig. Proc.*, 36:9, 1445-1453 (1988), an algorithm is  
13 described which solves the minimization problem. Specifically, the algorithm  
14 teaches that an unconstrained minimum of  $C_{\text{total}}(\mathbf{q}) + \lambda D_{\text{total}}(\mathbf{q})$  is also the  
15 solution to a constrained problem of the form required. The unconstrained  
16 problems are easier to solve; but the value of  $\lambda$  must be determined for the  
17 appropriate constrained problem. The constrained problem is thus  
18 transformed into a search through a family of unconstrained problems; and  
19 the algorithm of Shoham et al. gives appropriate bit allocations for the  
20 minimization problem to be solved by Section 18.

## 21 22 Channel Coding and Expected Image Distortion Section

23  
24 One objective of system 10 is to reduce or minimize the image distortion  
25 incurred in quantizing transform coefficients. Transmission of an image 14b  
26 over a network introduces a second source of distortion: network packet losses.  
27 In accord with the invention, the Channel Coding and Expected Image  
28 Distortion Section 36 controls quantization error by adaptively allocating  
29 quantizer resolution within the Bit Allocation for Source Coding Section 18

-21 -

1 via communication line 18a. In the same way, the Channel Coding and  
2 Expected Image Distortion Section 36 controls packet loss errors by selectively  
3 adding redundancy to the bitstream transmitted on the Internet 12. The image  
4 14a has already incurred loss during a lossy compression technique, e.g., such  
5 as through JPEG, and can generally withstand some additional loss during  
6 transmission, provided that those lost bits are not visually important. Because  
7 system 10 performs both source and channel coding jointly, the Channel  
8 Coding and Expected Image Distortion Section 36 knows the relative values of  
9 the bits within the image 14b, and thereby provides an extension of the above-  
10 described bit allocations by incorporating expected transmission losses into the  
11 distortion function and the costs of redundancy into the cost function.  
12 Specifically, the Channel Coding and Expected Image Distortion Section 36  
13 finds an optimized partition of bits into source and channel codes.

14  
15 The distortion variance can be controlled by adjusting the packet loss  
16 model used by the bit allocation algorithm. For example, by numerically  
17 increasing the assumed loss probability  $p_{loss}$ , the distortion variance can be  
18 controlled by adjusting the loss probability assumed in the optimizer. For  
19 example, numerically increasing the loss probability beyond the network's true  
20 packet loss rate has the effect of shifting bits from data to redundancy, which in  
21 turn increases the quantization distortion at a given bit-rate and also increases  
22 the protection against lost packets. Since the distortion variance functionally  
23 depends upon lost packets, increasing the degree of redundancy reduces the  
24 variance and increases image consistency.

25  
26 The problem thus addressed by the Channel Coding and Expected Image  
27 Distortion Section 36 is that of transmitting images as a collection of packets of  
28 bits of a maximum size  $S$  over the Internet 12 and network 32. The Channel  
29 Coding and Expected Image Distortion Section 36 has two separate properties

-22 -

1 for classes of network protocols: first, packets can be delivered out of order, so  
2 that each packet contains a unique identifier; and secondly, the contents of all  
3 packets are verified during transmission. Packets are generally lost for one of  
4 two reasons: a node somewhere on the Internet 12 and/or network 32 runs out  
5 of buffer space and drops the packet, or the packet is corrupted and fails a  
6 verification procedure somewhere in transit. Because of section 36's first  
7 property, i.e., that each packet contains a unique identifier, system 10 knows  
8 exactly which packets have been lost. Because of section 36's second property,  
9 i.e., that the contents of packets are verified during transmission, system 10  
10 assumes that all packets which are delivered are error-free because they have  
11 passed the protocol's verification procedure.

12  
13 To reduce decoded image variance and to facilitate the packing of  
14 subbands into equally sized network packets, e.g., such as 576 bytes each, the  
15 Channel Coding and Expected Image Distortion Section 36 breaks subbands (or  
16 subband bitplanes) into blocks of smaller memory sizes, each of which is  
17 preferably a maximum of 150 bytes. To reduce the visual impact of any losses,  
18 the Channel Coding and Expected Image Distortion Section 36 distributes pixels  
19 into these blocks through interleaving. All of the bitplanes of a subband are  
20 interleaved in the same way. In accord with the invention, blocks from a  
21 subband which represent different bitplanes, but which derive from the same  
22 image pixels, belong to the same interleaving.

23  
24 The Channel Coding and Expected Image Distortion Section 36 adds  
25 redundancy to the image transmission by adding FEC bits to the data stream.  
26 Because system 10 can tell which packets have been lost, a single block of FEC  
27 bits can protect a group of any number of blocks of data against single-packet  
28 loss. The Channel Coding and Expected Image Distortion Section 36 therefore  
29 conducts a tradeoff between protection and cost: greater protection of data is



-23 -

obtained by decreasing the size of the groups protected by FEC blocks, but increased protection increases the total transmission time because of the additional FEC blocks. The Channel Coding and Expected Image Distortion Section 36 thus performs this tradeoff in such a way so as to minimize the expected distortion of the image for a given total number of image bits, and, preferably, as a function of current network congestion, as described below.

In estimating the expected image distortion, the Bit Allocation for Source Coding Section 18 determines a probability of packet loss. To a first approximation, packet losses are independent Bernoulli trials, with losses occurring with probability P. However, studies of network traffic on a network such as network 32 reveal that network traffic is bursty and that these bursts are present across a wide range of time scales. See, W. E. Leland et al., On the self-similar nature of ethernet traffic, Proc. SIGCOMM, 183-193 San Francisco (1993). While it is true that very long bursts of losses in routers do occur, the rate adaptation that takes place in network protocols such as TCP greatly reduces the length of bursts actually experienced by the user. Accordingly, the Bit Allocation for Source Coding Section 18 incorporates bursts into a packet loss model such as through a first order Markov model. Specifically, the Bit Allocation for Source Coding Section 18 denotes a successful transmission by 0 and a loss by 1, and denotes the transition probabilities by  $P_{j,k}$ , where  $j, k \in \{0, 1\}$  correspond to the fates of two consecutive packets. The steady state loss rate is

$$P_1 = \frac{P_{0,1}}{P_{0,1} + P_{1,0}}$$

and the steady state success rate is  $P_0 = 1 - P_1$ . For consistency with the Bernoulli model,  $P_1 = P$ .

-24 -

1           To implement FEC according to one embodiment of the invention, data  
2 blocks are grouped into parity groups that are formed in of one of three ways by  
3 the Bit Allocation for Source Coding Section 18: (1) a parity group consisting of  
4 a single unshielded block; (2) a parity group consisting of multiple data blocks  
5 shielded by a single parity block; or (3) a parity block consisting of a single data  
6 block with multiple replicas. These three types of parity groups provide  
7 gradated levels of protection, ranging from minimal, unshielded blocks, to  
8 maximal, replicated blocks. For each subband, the Bit Allocation for Source  
9 Coding Section 18 determines a level of quantization refinement  $q_i$ , and a level  
10 of parity protection  $p_{i,k}$  for each subband bitplane.

11  
12           The use of the Markov model by section 18 enables the determination of  
13 the effects of burst errors within parity groups. A simplifying assumption is  
14 made that losses in different parity groups are independent. This between-  
15 group independence assumption is relevant only for parity groups containing  
16 blocks from the same subband. Section 18 also minimizes the effects of  
17 between-group correlation by interleaving the groups when loading packets  
18 with blocks.

19  
20           Section 36 also evaluates the effects of various levels of protection and  
21 quantization on coefficients in subband  $n$ . For illustration, let  $D_n$  be the average  
22 distortion resulting from setting a coefficient in subband  $n$  to zero. Let  $D_j$  be  
23 the average reduction in coefficient distortion given by bitplane  $j$ . When a  
24 high-order bitplane is lost, all lower order refinements will also be lost, since  
25 section 36 conditions the entropy coding of low order bitplanes on the high  
26 order values. Let  $X_i$  correspond to the event that the block containing bitplane  
27  $i$  for the coefficient is successfully transmitted. Section 36 then determines the  
28 expected distortion for the coefficient by the following relationships.

-25 -

1

2 Let  $d = \left\lceil \frac{M}{m} \right\rceil$  be the distance between successive blocks in an interleaved  
 3 FEC group. Interleaving the FEC group changes the effective success/loss  
 4 transition probabilities. We compute new transition probabilities  $P_{a,b}^{(d)}$  for the  
 5 interleaved blocks. Let  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$  be a binary string of length  $n$ . Under  
 6 our Markov model we have  $P(\sigma) = P_{\sigma_1} \prod_{k=1}^{n-1} P_{\sigma_k \sigma_{k+1}}$ . The new transition  
 7 probabilities are given by

8  $P_{a,b}^{(d)} = \sum_{k=0}^d \sum_{\sigma \in S_{d,k}^{(a,b)}} P(\sigma)$ . Let  $S_{n,k}^{(a,b)}$  denote the set of all length  $n$  binary strings that  
 9 begin with  $a$ , end with  $b$ , and contain exactly  $k$  1's. The probability of the string  
 10  $S_{n,k}^{(a,b)}$  can be determined recursively as follows. All probabilities  $P(S_{2,k}^{(a,b)})$  are  
 11 zero except  $P^{(d)}(S_{2,0}^{(0,0)}) = P_0 P_{0,0}^{(d)}$ ,  $P^{(d)}(S_{2,1}^{(1,0)}) = P_1 P_{1,0}^{(d)}$ ,  $P^{(d)}(S_{2,1}^{(0,1)}) = P_0 P_{0,1}^{(d)}$ , and  
 12  $P^{(d)}(S_{2,2}^{(1,1)}) = P_1 P_{1,1}^{(d)}$ . The probability of longer sequences is determined  
 13 recursively using the relation  $P^{(d)}(S_{n,k}^{(a,b)}) = \sum_{i=0}^k \sum_{(x,v) \in Q} P_a P^{(d)}(S_{n,i}^{(a,x)}) P_{x,v} P^{(d)}(S_{n,k-i}^{(v,b)})$ .

14 The probability  $P(X)$  of an unrecoverable loss of a block replicated  $m$  times with  
 15 spacing  $d$  is given by  $P(X) = P_1 (P_{1,1}^{(d)})^{m-1}$ . For a group of  $m$  blocks with spacing  $d$   
 16 that contains a single parity block, a block survives if there is at most one lost  
 17 packet. The probability that any block is lost unrecoverably is

18  $P(X) = \sum_{a=0,1} \sum_{b=0,1} \sum_{l=2}^m \frac{l}{m} P^{(d)}(S_{n,l}^{(a,b)})$ . In general, for an FEC scheme that can recover

19 from up to  $L$  losses, we have  $P(X) = \sum_{a=0,1} \sum_{b=0,1} \sum_{l=L+1}^m \frac{l}{m} P^{(d)}(S_{n,l}^{(a,b)})$ .

20

21 When loading blocks into network packets, section 36 imposes the  
 22 restriction that no two blocks from the same parity group may occupy the same

-26 -

1 packet so that the loss of one packet corresponds to the loss of only one  
 2 element of a parity group. An additional restriction is imposed to reduce the  
 3 variance of reconstructed images: no two blocks from different interleavings of  
 4 the same subband may occupy the same network packet or the same parity  
 5 group.

6  
 7 For the Bernoulli model,  $P(X) = 1 - P$  for unshielded blocks. When a  
 8 block is replicated, a block is lost only if all copies are lost. Hence,  $P(X) = 1 - P^m$ ,  
 9 where  $m$  is the total number of copies of the block that are transmitted. For  
 10 groups in which each  $m$  blocks are shielded by a single parity block,  $P(X) = 1 -$   
 11  $P[1 - (1 - P)^m]$ , which is one minus the probability of losing the given block and  
 12 at least one of the other  $m$  blocks in the group.

13  
 14 For the Markov model,  $P(X) = 1 - P = P_o$  for unshielded blocks. For  
 15 replicated and parity-shielded blocks, the order in which blocks are transmitted  
 16 to compute expected distortions must be determined. Accordingly, section 36  
 17 decorrelates losses within each parity group by spacing data as far apart as  
 18 possible in the transmitted bitstream, i.e., at packet intervals of  $\left\lceil \frac{M}{m} \right\rceil$  in an  $m$ -  
 19 block group, where  $M$  is the total number of packets to be transmitted as  
 20 determined from the total bit budget.

21  
 22 With  $P_{j,k}$  indicating the manner of parity shielding for bitplane  $k$  of  
 23 subband  $j$ , the Channel Coding and Expected Image Distortion Section 36  
 24 replaces the cost and distortion functions  $C_j(q_j)$  and  $D_j(q_j)$  with the functions  
 25  $C_{j,k}(q_j, p_{j,k})$  and  $D_{j,k}(q_j, p_{j,k})$  that incorporate the cost of the parity packets and the  
 26 expected distortion incurred in transmission. The new cost function  $C_j(q_j, p_{j,k})$   
 27 will equal the old  $C_j(q_j)$  plus the number of bits used for the parity blocks. The  
 28 new distortion function  $D_j(q_j, p_{j,k})$  is obtained using the expected distortion and

-27 -

1 success probabilities described above. As before, marginal analysis or Lagrange  
2 multipliers are preferably utilized for bit allocation, though with more choices  
3 for each iteration. That is, the Channel Coding and Expected Image Distortion  
4 Section 36 either increases the number of bit planes retained for a particular  
5 subband, or increases the parity protection for one particular subband or  
6 subband bitplane.

7  
8 Note that smaller groupings are more protective of the data and are thus  
9 less likely to distort the image with any packet losses. Because the invention  
10 preferably sorts by redundancy levels, Figure 1B illustrates typical data and  
11 redundancy groupings. In (a), three data blocks 100 are accompanied by one  
12 parity block 102. As such, only one of the data blocks 100 can be lost without  
13 image distortion. In (b), one data block 104 is accompanied with one parity  
14 block 106. In (c), one data block 108 is three-way replicated with two parity  
15 blocks 110. If the groupings are too small, such as in (b), the data 104 is  
16 essentially replicated. Each parity block 102, 106, 110 should be as long as the  
17 longest block in its parity group.

18  
19 Figure 1D illustrates a graphical distribution 115 of bitplanes and parities  
20 for one representative image (the Lena image discussed in detail in connection  
21 with Figures 4 and 5). In Figure 1D, black rectangles represent 64 byte data  
22 blocks and the gray rectangles represent the forward area correction (FEC) bits  
23 assigned to that data block. Leftmost blocks contain the highest order coefficient  
24 bitplanes; rightmost the lowest. The coarsest scale subbands are on the bottom  
25 117 of the chart 115; while the finest subbands are at the top 119. Note that the  
26 concentration of FEC bits is larger for higher order coefficient bitplanes and for  
27 coarser scale subbands.

28

-28 -

Those skilled in the art should appreciate that the expected distortion can be viewed and analyzed in other ways. For example, consider a subband consisting of  $n$  data blocks. Let  $D_q$  be the average quantization error incurred per block, let  $D_m$  be the error incurred in replacing all coefficients in a data block by the quantized subband mean, and let  $D_z$  be the error incurred in replacing all coefficients in a data block by zero. Since  $D_q \leq D_m \leq D_z$ , the zero-replacement is the worst-case scenario.

Data and parity blocks from each subband can therefore be distributed so that no two blocks from the same band are contained in the same network packet. Hence, losses of data blocks can be modeled as independent events. Every block transmitted, and every block that is lost and recovered produces an average error of  $D_q$ . If the subband mean is available, i.e., if at least one packet from the group is successfully transmitted, then the lost blocks produce an average error of  $D_m$ ; otherwise lost blocks produce an average error of  $D_z$ . The expected distortion for a band consisting of  $n$  data blocks is thus

$$E(D) = nD_q + np_{\text{unrecoverable}}(D_m - D_q) + np_{\text{unrecoverable}}^n(D_z - D_m),$$

where  $p_{\text{unrecoverable}}$  is the probability of an unrecoverable packet loss.

### Encoding and Decoding

With reference to Figure 2, the steps for encoding and decoding transmitted images between the system 10 and any of the users 30, 34a, 34b are as follows. The encoding occurs in the system 10, prior to transmission on the Internet, and the decoding takes place after transmission and within any of the decoding sections 35.

1  
2 More particularly, system 10 encodes images by the following steps.  
3 First, in step 40, the Image Compression Section 16 applies a compression  
4 algorithm, e.g., a wavelet subband decomposition, to the image 14 to form a  
5 compressed image representation 14a. Thereafter, in step 42, the Bit Allocation  
6 for Source Coding Section 18 decomposes each subband in image 14a into a  
7 series of nested quantizers, i.e., bitplanes, and, in step 44, determines which  
8 bitplanes to send, and how much redundancy to assign to each bitplane. The  
9 resulting image file from Section 18 is shown as image 14b.

10  
11 In step 46, the Channel Coding and Expected Image Distortion Section  
12 36 then interleaves the bitplanes into blocks of some fixed size, i.e., the size  
13 after compression. For example, blocks of 150 bytes each can be used. Even  
14 though smaller blocks could be formed to reduce image variance, by spreading  
15 losses more evenly, smaller blocks also introduce overhead. Thus, interleaving  
16 visually decorrelates losses, although it has no effect on system 10's quality  
17 metric. Interleaving has little or no effect on compression performance  
18 because the interleaving occurs after the subband decomposition.

19  
20 In step 48, the Channel Coding and Expected Image Distortion Section  
21 36 compresses the formed blocks using an arithmetic coder. Nested  
22 quantization can be coded in the same total number of bits as a non-nested  
23 quantization of the same bins by (i) coding in decreasing bitplane magnitude  
24 order, and by (ii) using the high order bits for a transformed pixel as a context  
25 selecting frequency tables for low order bits.

26  
27 Finally, in step 50, the Channel Coding and Expected Image Distortion  
28 Section 36 adds redundancy to the image and packs blocks into network  
29 packets. Blocks of the same protection level are grouped into parity groups of

-30 -

1 appropriate size, such as shown in Figure 1B. For example, 576 bytes per  
2 network packet is the current maximum size of unfragmented Internet  
3 transmissions. Accordingly, 576-byte network packet sizes are preferably used  
4 with the invention; though those skilled in the art should appreciate that  
5 other packet sizes can be used, particularly if Internet protocol changes. Sorting  
6 by size is useful because the parity block is as large as the largest data block in  
7 the group. Blocks in the same FEC group are preferably spaced out as much as  
8 possible to take advantage of burst losses.

9  
10 With respect to Figure 2A, each of the decoding sections 35 operate to  
11 reverse the above-described encoding steps, except that the bit allocation  
12 decisions have already been made, and some of the packets may have been lost  
13 during transit. In particular, in steps 52 and 53, the decoding section 35 first  
14 reads the surviving packets and sorts those packets into parity groups. If any  
15 parity groups have one or missing members, section 35 reconstructs the  
16 missing member in step 54.

17  
18 In step 55, each decoding section 35 decodes all of the data blocks into  
19 their respective subband bitplanes. If a high-order bitplane block is missing,  
20 then all of the lower order bitplane blocks corresponding to the high order  
21 block are not decodable. Finally, in step 56, the image is reconstructed: missing  
22 coarse band pixels are reconstructed by averaging neighbors, while missing  
23 detail band pixels are reconstructed using the subband mean. In Figure 2A, the  
24 reconstructed image 14c is shown on computer display 60 of client station 62  
25 with image features as shown in image 14 of Figure 1.

26  
27 In a flow chart format, Figure 2B illustrates alternative encoding  
28 methodology according to the invention. In step 40', a wavelet subband  
29 decomposition is applied to the image 14'. In a complete wavelet



-31 -

1 decomposition, the coarse-scale subband is a single pixel value corresponding  
2 to a weighted average of all pixels in the image 14'. Because certain header  
3 information is maintained with each subband, it is more efficient to stop the  
4 wavelet transformation at some point short of a single pixel, e.g., a 32 x 32  
5 coarse-scale image, and to transmit this image untransformed. This base image  
6 is referred to herein as the "coarse scale subband," similar to the DC band of a  
7 JPEG image. The detail (non-DC) bands are thus refinements of the image,  
8 whereby each successive band provides the information necessary to double  
9 the image resolution. Image 14a' illustrates the decomposition of image 14'.

10  
11 In step 42', quantizer redundancies and parity levels are assigned, for  
12 example as described in connection with the Bit Allocation for Source Coding  
13 Sections 18, 18' of Figures 1 and 1A.

14  
15 In step 44', each band is distributed across many packets by interleaving  
16 pixels so that a lost packet will not cause a catastrophic band loss. For example,  
17 Turner et al., Image transfer: an end-to-end design, SigComm 92, 258-268,  
18 describes one such suitable interleaving scheme. Distributing subbands does  
19 not reduce expected distortion because the chance of some loss in a given  
20 subband is increased by distribution; but it does reduce the variance in the  
21 expected distortion by increasing the population subject to the transmission  
22 experiment.

23  
24 The incentive for subdivision is tempered by the desired step of  
25 encoding a descriptive header within each independent image block. Since  
26 image transmission is lossy, and since only an unknown subset of network  
27 packets arrive intact, a header which describes enough detail so as to permit  
28 lossy reconstruction of the block's subband is added to each block. For example,  
29 an image block of up to about 150 bytes provides suitable subdivision.

1 Likewise, header information can be about 15 bytes per block, so that if many  
2 small subbands are not broken into blocks, roughly 10% of the compressed  
3 image ends up as header information.

4  
5 Alternatively, those skilled in the art should appreciate that the header  
6 could be located in a few heavily shielded packets, providing a more efficient  
7 configuration since much of the header information is replicated between  
8 blocks from the same subband.

9  
10 In step 46', the wavelet coefficients are compressed within the relevant  
11 block. In one embodiment of the invention, these wavelet coefficients are  
12 compressed using adaptive arithmetic coding. Arithmetic coders emit  $\Sigma - \log_2$   
13  $p_i$  bits where  $p_i$  is the predicted probability of the  $i$ th event. In adaptive  
14 coding, the relative frequencies of past events are remembered in histograms  
15 and are used to estimate the probability of future events for the purposes of  
16 coding. To ensure that no event has a predicted probability of zero, histograms  
17 are usually initialized so that all possible events have a frequency of one. The  
18 histogram is adapted to the actual frequencies encountered as the input is read.  
19 For a large dataset, the inertia represented by the initial flat histogram is  
20 relatively unimportant.

21  
22 The amount of time available for the histogram to adapt to the input  
23 distribution is reduced through subdivision into blocks. To compensate for  
24 this effect, the FLIIT methodology of this embodiment preferably incorporates  
25 step 48'. Specifically, step 48' utilizes the following two-histogram scheme,  
26 which adapts much more quickly than a single histogram scheme:

- 27  
28 ● Initialize two histograms, one histogram (F) that is flat with every  
29 possible value initialized to one, and one histogram (H) that is empty

-33 -

1 with a single symbol, e.g., the escape symbol with an initial probability  
2 and frequency of one.

3  
4 ● Whenever an input symbol appears with non-zero probability  
5 (frequency) in histogram H, code the input symbol within histogram H  
6 and increment its frequency therein.

7  
8 ● Whenever an input symbol appears with zero probability in  
9 histogram H, code the escape symbol within histogram H, and code the  
10 input symbol within histogram F. This new symbol is added to  
11 histogram H with a frequency of one, and histogram F is never again  
12 used to code this symbol.

13  
14 In step 50', after the compressed blocks are generated, redundancy is  
15 added. First, the blocks are sorted by protection level and size, in order of  
16 decreasing protection and size. Blocks requiring replication are replicated; and  
17 a given block and its replicas are all assigned the same parity group number,  
18 which prevents them from being included in the same network packet.

19  
20 Blocks requiring the same level of parity protection are preferably  
21 grouped together by type, even though it is not generally possible to meet this  
22 preference exactly. For example, four data blocks can exist which are preferably  
23 arranged in a group of five data blocks protected by a parity block. In such a  
24 case, one block with less stringent protection requirements is promoted, if  
25 available, to round out the group. The promotion of this block is more  
26 efficient than the alternative, which leaves parity groups unfilled, and which  
27 effectively promotes all of the members of a group to a higher level of  
28 protection.

-34 -

1  
2         Sorting by size also helps to keep similarly-sized blocks in the parity  
3 group, which is valuable because the parity block should be as large as the  
4 largest data block in the group.

5  
6         In step 51', the data and parity blocks are readied for transmission.  
7 When the network 12' is congested, throughput will be gated by router  
8 scheduling, rather than by bandwidth. Routers schedule communication  
9 channels using a round-robin-type algorithm which is insensitive to packet  
10 size. Accordingly, users of packets which are smaller than the largest packet  
11 transmitted by the network will pay a throughput penalty. Conversely, users  
12 who send overly-large packets that become fragmented en-route lose an entire  
13 packet whenever a single fragment is lost, also resulting in reduced  
14 throughput.

15  
16         Because the largest Internet packet size which is guaranteed transmission  
17 on the Internet 12' without fragmentation is 576 bytes, the invention preferably  
18 groups information into packet sizes which are 576 bytes. Those skilled in the  
19 art should appreciate that differing network packet sizes can be used with the  
20 invention and that the 576-byte network packet size is subject to change with  
21 the growth and expected protocol changes of the Internet 12'. By way of  
22 example, one suitable packet packaging includes a largest-first first-fit heuristic  
23 protocol to pack blocks into 550-byte UDP packets. ATM further incorporates  
24 data blocks of 58 bytes; and this size is also suitable for the invention. If desired,  
25 the protocols of the invention can include additional restrictions: that blocks  
26 from the same parity group are not allowed in the same packet, and that blocks  
27 from the same band are not allowed in the same packet.

28

1 Rate Control Section

2  
3 Because the Internet is a shared medium, system 10 of Figure 1  
4 preferably controls the rate at which data is transmitted thereon or risks  
5 causing congestion in the network, resulting in lost packets and reduced  
6 performance for every connected user. In the prior art, TCP protocol  
7 implemented in the Reno release of BSD UNIX, known to those skilled in the  
8 art, controls its rate by starting out very slowly, by slowing down when packets  
9 are dropped, indicating congestion, and by speeding up otherwise. The  
10 problem with the TCP Reno strategy is that it simultaneously induces a certain  
11 level of packet losses on the Internet.

12  
13 A second prior art rate control method implemented in TCP Vegas, see,  
14 e.g., L. S. Brakmo et al., TCP Vegas: New techniques for congestion detection  
15 and avoidance, Proceedings of the SIGCOMM '94 Symposium (1994), operates  
16 to compare expected throughput rates with actual throughput rates.  
17 Whenever the rate of packet reception drops below the rate of packet  
18 transmission, the network must be storing or dropping the excess data.  
19 Accordingly, TCP Vegas does not require packet losses in order to function, and  
20 typically delivers higher throughput than TCP without degrading the  
21 performance of other TCP connections.

22  
23 Neither of the prior art TCP Reno or TCP Vegas rate control schemes are  
24 appropriate for the transmission of compressed images across the Internet.  
25 They are not appropriate because compressed images are much smaller than  
26 the dataset size required to achieve steady state transmission. By way of  
27 example, because TCP Reno overshoots the channel's actual throughput by a  
28 factor of two at the end of a slow start-up, a delay of 2.5 seconds before steady  
29 state can be measured as well as heavy packet losses at the end of start-up. See

1 L. S. Brakmo et al., TCP Vegas: New techniques for congestion detection and  
2 avoidance, Proceedings of the SIGCOMM '94 Symposium (1994).

3  
4 In accord with the invention, FLIIT methodology preferably implements  
5 rate control using a delay-based scheme, whereby FLIIT clients, e.g., any of the  
6 users 30, 34a, 34b, sends an acknowledgment along Internet 12 to the Rate  
7 Control Section 22 every 16th packet. Rate Control Section 22 uses this  
8 acknowledgment to measure the round trip time ("RTT") and the current loss  
9 rate. The smallest observed RTT is the base round trip time ("BRTT"), which  
10 is assumed to be the round trip time in an uncongested network and which is  
11 generally the fastest round trip travel time. System 10 attempts to keep the  
12 actual RTT just above the BRTT by adjusting the sending rate. In other words,  
13 system 10 attempts to keep a small constant number of packets stored in the  
14 network which are ready in case congestion drops; yet there are not so many  
15 stored packets that they contribute to losses.

16  
17 On the Internet, the number of connected networks relates to the  
18 "packet store," which is the number of packets in the Internet routers. Figure  
19 1C shows a typical transmission of a packet 114 through various Internet  
20 routers 112. The routers 112 act as FIFOs because the first packet 116 within each  
21 packet store 118 is transmitted to the next location, which could be a another  
22 router or the end recipient. One goal of the Rate Control Section, therefore, is  
23 to ensure that some packets, but not too many, are transmitted to and stored  
24 within the routers 112. It adjusts the transmission rate so as to select the  
25 number of packets within the routers, to maximize image quality as a function  
26 of time.

27  
28 The packet store  $S$  corresponding to a given RTT is thus the number of  
29 packets sent during that interval. The extra packet store  $\Delta S = S (\text{RTT} -$

-37 -

1 BRTT)/RTT, and system 10 operates to keep  $\Delta S$  in the range  $\Delta S_L < \Delta S < \Delta S_H$ ,  
 2 where  $\Delta S_H$  and  $\Delta S_L$  are determined empirically.  $\Delta S_H$  is the respective router's  
 3 estimate of how large the packet store can be above the first packet, i.e., those  
 4 packets in the store 118 above the packet 116, Figure 1C.  $\Delta S_H$  should be set low  
 5 enough so that system 10 does not over-drive the network; and  $\Delta S_L$  should be  
 6 set sufficiently high so that system 10 responds to available network  
 7 bandwidth.

8  
 9 Given  $S$ ,  $\Delta S_H$ , and  $\Delta S_L$ , system 10 adjusts the interpacket sending interval  
 10  $I = \text{RTT}/S$  using a method such as Newton's iterative method. If, for example,  
 11  $\Delta S < \Delta S_L$ , system 10 can increase the sending rate to  $I_{\text{new}} = I - I(\Delta S_L - \Delta S)/S$ . If, on  
 12 the other hand,  $\Delta S > \Delta S_H$ , system 10 can decrease the sending rate to  $I_{\text{new}} = I + I$   
 13  $(\Delta S - \Delta S_H)/S$ .

14  
 15 System 10 can also function to react to systematic packet losses, while  
 16 ignoring sporadic packet losses. By way of example, if losses over some  
 17 threshold  $T_1$ , between acknowledgments, cause a small reduction in the  
 18 sending rate,  $I_{\text{new}}$  can be set to  $I_{\text{old}} F$ , where  $F$  is a constant less than one (e.g.,  
 19 0.9), to slow down the response. If on the other hand, losses over a larger  
 20 threshold  $L > T_2$  cause a larger reduction in the sending rate, then  $I_{\text{new}}$  can be  
 21 set to  $I_{\text{old}} f$ , where  $f$  is a constant smaller than  $F$  (e.g., 0.5).

22  
 23 Those skilled in the art should appreciate that other rate control  
 24 schemes can be used with the invention. For example, an off-line process can  
 25 be used to select a transmission rate for FLIIT packets, such as by picking the  
 26 knee on the network's load/loss curve. See, e.g., C.L. Williamson et al., Loss-  
 27 load curves: Support for rate-based congestion control in high-speed datagram  
 28 networks, Proceedings of SIGCOMM 91, pp. 17-287 (1991). In accord with the  
 29 invention, streams of packets containing roughly 550 bytes were sent at various

-38 -

1 transmission rates, and the loss rate was measured for each rate. As shown in  
2 Figure 2C, the loss rate as a function of transmission rate was relatively  
3 constant at rates below about 4ms-per-packet, as compared to higher rates. Each  
4 curve of Figure 2C corresponds to a different transmission rate. Above 4ms-  
5 per-packet, however, the loss rate increased sharply, relative to rates  
6 corresponding to 4ms-per-packet and below. To avoid high loss rates, and to  
7 avoid impacting other applications operating on the Internet, a reasonable  
8 transmission packet rate is therefore 4ms-per-packet, in accord with the  
9 invention. Faster transmission rates can be chosen at other times. For  
10 example, rates of 2ms-per-packet are effective between about 03:00 and 04:00  
11 EDT; but such rates generate high losses when the Internet becomes busy  
12 during the day. During daylight hours, the loss rate never drops much below  
13 about 5% regardless of the transmission rate.

14  
15 The rate control section of the invention thus provides certain  
16 advantages over the art. For example, because the buffer capacity of the Internet  
17 between any two well-separated nodes is typically greater than the size of a  
18 well-compressed image, a server could transmit an entire image in less than  
19 one RTT. This however is not feasible with TCP because TCP has a slow start-  
20 up time for each connection; and further takes many round trips to reach full  
21 speed. This is fine for megabyte transfers, but inappropriate for smaller and  
22 widely used image sizes, e.g., 8-kilobyte images. In a preferred embodiment of  
23 the invention, therefore, the FLIIT server remembers the effective transfer  
24 rates across its active connections, and effectively removes the slow startup as  
25 an issue



## Stopping Criterion

In certain instances, FLIIT packets may be lost or delayed for long periods of time. If system 10 (Figure 1) waits too long for slow packets, system 10 loses responsiveness. If, on the other hand, system 10 does not wait long enough, it loses packet data. This tradeoff, between transmission speed and packet loss, has a practical solution, in accord with the invention. Specifically, in a preferred embodiment, system 10 incorporates the tradeoff into the resource allocation algorithm to choose an optimal time to stop waiting for lost packets. Typically, the ideal stopping point is the expected time of arrival of the last packet plus the standard deviation of the interpacket arrival time.

Other stopping criterion can be used with the invention. For example, in one embodiment of the invention, system 10 of Figure 1 sends packets at a constant rate, which preferably keeps Internet congestion down. If, for example, a packet is sent every  $b$  time units, where  $b$  is less than or equal to the throughput of the network, and the network delivers all packets after a fixed time delay, then the  $n$ -th packet will arrive at time  $T_n = a + (n - 1)b$ , where  $a$  is the arrival time of the first packet. On the Internet 32, packets are delayed by variable lengths of time. System 10 can incorporate this variability into the arrival time model by adding a random delay variable  $X_n$ . Accordingly,  $T_n = a + b(n - 1) + X_n$ . System 10 then determines a stopping time  $T_{\text{stop}}$  after which it stops waiting for packets to reconstruct the image 14. Packets arriving after time  $T_{\text{stop}}$  are then considered lost by system 10.

Given the distributions of  $X_n$ , the probability  $P(T_n > T_{\text{stop}})$  that packet  $n$  will be lost due to excessive delay can be determined. By randomizing the order of the packets transmitted, the probability of any given packet being the  $n$ -th

-40 -

1 packet transmitted is  $1/N$ , where  $N$  is the total number of packets sent. The  
 2 probability of a particular packet being lost due to delay is

$$P_{\text{delay}}(T_{\text{stop}}) = \frac{1}{N} \sum_{k=1}^N P(T_n > T_{\text{stop}})$$

3  
 4  
 5  
 6 The overall probability that a packet is lost is then  $p_{\text{loss}} = 1 - (1 - p_{\text{drop}})(1 - p_{\text{delay}})$ ,  
 7 where  $p_{\text{drop}}$  is the probability of the packet being dropped in transit.

8  
 9 The stopping time affects the loss rate observed by the receiver, e.g., any  
 10 of the units 34a on the Internet 32. The reconstructed image distortion is thus a  
 11 function of the number of data, the redundancy bits, and the stopping time.  
 12 Because the constraint is on the number of bits sent, and not on the length of  
 13 time required to receive the image, the optimal value of  $T_{\text{stop}}$  is infinity. If the  
 14 goal is to maximize responsiveness, the time required to receive the image is  
 15 constrained rather than the total number of bits sent. This can be done by  
 16 setting the cost function as the sum of the time required to send the bits in the  
 17 image plus the waiting time. This results in a new set of cost and distortion  
 18 functions which depend on the bit allocations as well as the stopping time. By  
 19 varying the stopping time in the allocation algorithm, discussed above, the bit  
 20 allocations and stopping time can be obtained in an optimized fashion.

21  
 22 Figure 2D illustrates observed and fitted cumulative density functions  
 23 for the packet delays  $X_n$ , which correspond to a set of independent, identically  
 24 distributed Poisson random variables with parameter  $\lambda$ . The data was gathered  
 25 from ten 160-packet transmissions. Through the method of moments, the  
 26 offset rate,  $a$ , and the sending rate,  $b$ , can be determined by least squares and the  
 27 parameter  $\lambda$  to isolate the delay  $X_n$ . The resulting delay is normalized to have

-41 -

1 mean 0 and variance 1. The superimposed solid curve is the cumulative  
2 density function for an equivalently normalized Poisson random variable.

3  
4 As illustrated in Figure 2D, the stopping time model of this embodiment  
5 describes the distribution of delays. The server can update its knowledge of  
6 network conditions by periodically obtaining these quantities from the  
7 receiver. The typical stopping time is the expected time of arrival of the last  
8 packet,  $a + b(N - 1) + \lambda$ , plus a delay ranging from 0 to the square root of  $\lambda$ ,  
9 which is the standard deviation of the delay.

## 10 11 Experimental Results

### 12 13 (1) Experiment 1.

14  
15 Using the well known Lena image at 256 x 256 resolution, we generated  
16 sets of packets using the FLIIT methodology discussed herein, as well as three  
17 different fixed-parity schemes, such as shown in Figure 3. The fixed-parity  
18 schemes used the same bit allocation as in FLIIT in order to determine  
19 quantizer resolutions for each subband, but no adaptive-coding was done for  
20 the parity bits. Experiment 1 therefore shows only the effects of adaptive  
21 versus fixed distribution of redundancy. In the fixed parity 3 scheme, each data  
22 block was replicated three times. In the fixed parity 1/3 scheme, groups of three  
23 data blocks were protected by a single parity block. In the fixed parity 0 scheme,  
24 no parity blocks (redundancy) were used. The "Y" axis of Figure 3 represents  
25 the peak signal to noise ration (PSNR), a logarithmic indicator of image  
26 quality; and the "X" axis represents the expected loss rate. As shown, the FLIIT  
27 methodology of the invention dominates the other schemes, usually by  
28 several dB.

1  
2 In Experiment 1, packets were generated within each scheme using 8:1  
3 compression, and with expected loss rates ranging from 0% to 50%. For each  
4 combination of parity scheme, compression ratio, and loss rate, we ran  
5 simulated transmission experiments in which packets were deleted by  
6 subjecting each packet to an independent pseudo-random Bernoulli trial.  
7 Images were then reconstructed from the remaining packets, allowing image  
8 comparisons and calculations of actual image distortions.  
9

10 Figures 4A, 4B, 4C illustrate the fixed parity transmission results of  
11 Experiment 1. Figures 5A, 5B and 5C illustrate Lena image transmissions  
12 under FLIIT testing of Experiment 1. Compared to the loss rates tested, the  
13 FLIIT methodology of the invention has the overall best performance. The  
14 fixed parity 3 scheme performs best for high loss rates because of the large  
15 amounts of transmitted redundancy. At high loss rates, FLIIT also uses large  
16 amounts of redundancy, but it distributes these redundancy bits more  
17 selectively than the fixed scheme. In particular, FLIIT methodology shields the  
18 low-frequency portions of the image since the loss of a low frequency data block  
19 results in a much larger error than the loss of a high frequency block. The extra  
20 shielding is also relatively inexpensive, since there are relatively few low  
21 frequency coefficients.  
22

23 More particularly, Figures 4A-4C and 5A-5C show, respectively, the  
24 effects of compression and transmission losses on the 256 x 256 Lena image  
25 under the fixed parity 3 scheme and under FLIIT. The images have been  
26 compressed from 64K to 9.5K (8:1 compression plus a roughly 20% packet  
27 header overhead cost), including the parity blocks, and all packets have a 50%  
28 probability of being lost. In effect, these images have been reconstructed from  
29 4K of randomly selected data. These data show that transmissions with FLIIT

-43 -

1 methodology perform well even at very high error rates. Figures 4A and 5A  
2 have a 90th percentile image quality; Figures 4B and 5B have a 50th percentile  
3 image quality; and Figures 4C and 5C have a 10th percentile reconstructed  
4 image quality. These figures represent a very severe test as images are reduced  
5 to roughly 9Kbytes (18-20 packets with overhead) and then packets are  
6 randomly eliminated in independent trials, so that well under 50% of the  
7 packets typically survive.

8  
9 Figure 5D illustrates the Lena image of Figures 4 and 5 transmitted via  
10 TCP/IP. In comparison, the FLIIT transmitted images of Figures 5A-5C  
11 required between 0.8 to 2.0 seconds to transmit, while the TCP/IP transmitted  
12 image typically required between 1.4 and 12.3 seconds to transmit. On average,  
13 11.5% of Internet packets were lost during each transmission of Figures 5A-5C.  
14 FLIIT thus trades variability in image delivery time for variability in image  
15 quality.

16  
17 (2) Experiment 2.

18  
19 In Experiment 2, we measured image quality (PSNR) as a function of  
20 transmission time for both FLIIT and TCP. Time begins when a client requests  
21 an image, and ends when the client decides that it has received an image. We  
22 did not include the decode time, which is the same for both clients, and which  
23 is practically negligible. For FLIIT protocol, the client makes its request with a  
24 single UDP packet. For TCP transport, the client makes its request over a TCP  
25 connection. FLIIT images are returned using UDP. TCP images are returned  
26 using TCP. The TCP images were generated using the same compression  
27 routines as the FLIIT images, but there was no redundancy or blocking,  
28 eliminating all of the over-head which FLIIT needs for reconstruction after  
29 lossy transmission, but which are unnecessary after lossless transmission.

-44 -

1 Accordingly, the prior art TCP method was not burdened or handicapped with  
2 the overhead byte requirements of FLIIT.

3  
4 In accord with the invention, discussed above, the FLIIT client calculates  
5 a running estimate of the expected time of arrival of the last packet. The client  
6 waits some period beyond this time, typically one standard deviation of the  
7 interpacket arrival time, and decodes the image. The exact amount of extra  
8 time to wait is calculated and specified by the FLIIT server. The TCP client stops  
9 when the complete image has arrived.

10  
11 We used a real Internet connection for this experiment. The connection  
12 was between Dartmouth College in Hanover New Hampshire, and Stanford  
13 University in Stanford California. The participating computers were separated  
14 by 20 hops. For convenience, we ran the client and the server locally, but sent  
15 the data across the continent, by routing network packets from our local client  
16 to a local pseudo server, which bounced these packets off of the Stanford  
17 machine's echo server, and forwarding the returning packets to our local  
18 server. Traffic from the server to the client was also similarly redirected  
19 through the remote echo server. The transport methodologies used during  
20 experimentation are illustrated in Figure 6. Data packets for FLIIT methodology  
21 originated from a local FLIIT user 70 and through the forwarding server 72 at  
22 Dartmouth College. Stream data for TCP originated at a local TCP user 74 and  
23 similarly forwarded through server 72. The data from either transport  
24 methodology was thereafter forwarded through an echo server 76 at Stanford  
25 University, and then back to the respective servers 70, 74.

26  
27 The image used in Experiment 2 was again the Lena image at 256 x 256  
28 resolution. We transmitted Lena at different compression ratios, 160 times for  
29 each sample. We ran the experiment under two different sets of circumstances:

-45 -

1 daytime and nighttime, both on weekdays. Daytime was 12:00-18:00EDT.  
2 Nighttime was 02:00-08:00EDT. We set the expected loss rate, expected packet  
3 arrival rate, and standard deviation of interpacket delay to 1.3%, 4.4ms per  
4 packet, and 10.4ms for the night experiments, and 8.2%, 4.6ms per packet, and  
5 12.3ms during the day.

6  
7 The results of this experiment are graphed in Figure 7, which plots  
8 image quality in terms of PSNR, a logarithmic function of the mean squared  
9 error, and as a function of transmission time. Plotted points correspond to  
10 median values, while error bars indicate first and third quartiles. TCP curves  
11 have error bars only in the time dimension because they deliver consistent  
12 quality. FLIIT has error bars in both dimensions because both quality and time  
13 are variable. For equivalent quality, FLIIT methodology is roughly twice as fast  
14 as TCP at night, and four times faster than TCP during the day. FLIIT has  
15 minimal variation in transmission time, while TCP transmission times vary  
16 widely, especially during the day.

17  
18 FLIIT methodology, in accord with the invention, uniformly  
19 outperformed TCP for equivalent image quality. Highly compressed FLIIT  
20 images were transmitted over twice as fast as their TCP counterparts,  
21 presumably because fewer round trips are necessary to establish a FLIIT  
22 connection. Moderately compressed images were transmitted more quickly  
23 because FLIIT does not retransmit dropped packets or wait multiple round trip  
24 times for the last few packets. During the day, when the Internet is congested,  
25 FLIIT methodology is more than four times faster than TCP, even for high  
26 quality images.

27  
28 In summary of Experiment 2, FLIIT accepts some variance in quality for  
29 a large improvement in throughput and a large reduction of the multisecond

1 variance in time accepted by TCP. Although TCP makes the right tradeoff for  
2 applications requiring perfect transmission, FLIIT methodology, in accord with  
3 the invention, makes the right tradeoff for interactive and real-time  
4 applications. Other experimentation with wavelet-based coder transforms,  
5 according to the invention, has yielded PSNR's for the 512 x 512 Lena image of  
6 image within 0.3 to 0.9 dB of images created by very high quality prior art  
7 coders such as described by J.D. Villasenor et al., IEEE Trans. Image Processing  
8 (1995).

9  
10 (3) Experiment 3.

11  
12 In a third experiment, a comparison was made between the impact of a  
13 FLIIT operation on a TCP session and the impact of a TCP operation on a TCP  
14 session to determine whether FLIIT methodology over-utilizes the Internet as  
15 compared to TCP. Because of a possible concern that some of FLIIT's  
16 performance might derive at the expense of other network clients, e.g., that  
17 FLIIT methodology might appropriate disproportionate bandwidth away from  
18 TCP connections, Experiment 3 was conceived to measure, separately, the  
19 performance of (1) FLIIT alone, (2) TCP alone, (3) FLIIT with FLIIT, (4) TCP with  
20 TCP, and (5) FLIIT with TCP. The format of the experiment was otherwise the  
21 same as the Experiment 1, e.g., same image, same network, same number of  
22 trials, etc.

23  
24 In the paired methods of Experiment 3, two images were transferred  
25 together, and were allowed to finish separately. Because FLIIT image transfers  
26 always finish much sooner than TCP, the TCP vs. FLIIT experiments show the  
27 performance of TCP running by itself for much of the time. This experiment  
28 particularly measures the impact on the network of FLIIT vs. TCP protocols  
29 performing the same task.



-47 -

1  
2        Figures 8A, 8B illustrate the results of Experiment 3. Each graph plots  
3 image quality as a function of time. Figure 8A shows the performance of FLIIT  
4 by itself, FLIIT competing for bandwidth with a TCP connection, and FLIIT  
5 competing for bandwidth with a FLIIT connection. In Figure 8B, analogous  
6 graphs feature TCP performances. Because Experiment 3 was carried out on  
7 the Internet, which has many other users, there is no expectation that two  
8 simultaneous transmissions should take twice as long as one transmission.  
9

10        Even though FLIIT transmission were generally faster, the two FLIIT  
11 connections of Figure 8A degraded each other's performance more than the  
12 two TCP connections of Figure 8B. This indicates that FLIIT utilizes a larger  
13 fraction of the aggregate network bandwidth than TCP, so that when two FLIIT  
14 connections run together, they have a larger effect on each other.  
15

16        For low quality image transmissions, TCP runs about as fast competing  
17 with FLIIT as it does competing with TCP. For medium quality images, TCP  
18 runs slightly slower with FLIIT as compared to running with TCP. At the  
19 highest quality, TCP runs faster with FLIIT than with TCP. This indicates that  
20 transferring a high quality image with FLIIT has less effect on the network than  
21 transferring a high quality image with TCP, but that while FLIIT is transferring  
22 an image, it has a greater effect on the network than does TCP.  
23

24        The invention thus demonstrates a system which combines source and  
25 channel coding to produce an image transfer protocol that transfers images of a  
26 given quality twice as fast as the TCP protocol at night, and four times faster  
27 than TCP during the day. Note that this figure is comparing wavelets to  
28 wavelets. Further, FLIIT will outperform JPEG image transmission by an even  
29 greater margin, since JPEG images are larger than wavelet images.

-48 -

1  
2       The FLIIT methodology presented herein is particularly appropriate for  
3 image previewing, progressive image transmission, transmission of moving  
4 pictures, and broadcast applications.

5  
6       Those skilled in the art should appreciate that changes can be made  
7 within the description above without departing from the scope of the  
8 invention. For example, it should be apparent that image compression within  
9 the Image Compression Section 16, Figure 1, can utilize DCT-based schemes  
10 such as JPEG by replacing wavelet subbands, described above, with blocks of  
11 DCT coefficients of comparable frequencies.

12  
13  
14  
15  
16  
17  
18       The invention thus attains the objects set forth above, among those  
19 apparent from preceding description. Since certain changes may be made in  
20 the above apparatus and methods without departing from the scope of the  
21 invention, it is intended that all matter contained in the above description or  
22 shown in the accompanying drawing be interpreted as illustrative and not in a  
23 limiting sense. It is also to be understood that the following claims are to cover  
24 all generic and specific features of the invention described herein, and all  
25 statements of the scope of the invention which, as a matter of language, might  
26 be said to fall there between.

27  
28       Having described the invention, what is claimed as new and secured by  
29 Letters Patent is:

-49 -

1     1.     A method of decomposing a digitized signal into a collection of  
2     subbands for transmission over the Internet, comprising the steps of (a)  
3     allocating forward error correction bits and quantizer precision to each subband  
4     in order to reduce expected image distortion subject to an overall bit budget;  
5     and (b) transmitting the quantized image and forward error correction bits over  
6     the Internet.

7  
8     2.     A method according to claim 1, wherein the step of allocating forward  
9     error correction bits and quantizer precision to each subband comprises the step  
10    of minimizing expected image distortion subject to the bit budget.

11  
12    3.     A method according to claim 1, wherein the step of allocating forward  
13    error correction bits and quantizer precision to each subband occurs  
14    automatically.

15  
16    4.     A method according to claim 1, wherein the step of allocating forward  
17    error correction bits comprises the step of decomposing the subband into  
18    outputs of quantizers.

19  
20    5.     A method according to claim 4, wherein the step of allocating forward  
21    error correction bits comprises the step of allocating forward error correction  
22    bits to each of the outputs.

23  
24    6.     A method according to claim 4, wherein the step of decomposing the  
25    subband into outputs of quantizers comprises decomposing the subband into  
26    outputs of nested quantizers.

27

-50 -

1 7. A method according to claim 6, wherein one or more of the outputs  
2 comprise a succession of discrete refinements to a discrete representation of a  
3 real value.

4  
5 8. A method according to claim 6, comprising the further step of entropy  
6 coding the outputs.

7  
8 9. A method according to claim 8, wherein the entropy coding is selected  
9 from the group of adaptive arithmetic coders, static arithmetic coders,  
10 Huffman coders, and dictionary-based coders.

11  
12 10. A method according to claim 9, wherein the dictionary based coders  
13 comprise one of Lempel-Ziv 77 coder and Lempel-Ziv 78 coder.

14  
15 11. A method according to claim 4, wherein one or more outputs comprise  
16 a discrete representation of a real value.

17  
18 12. A method according to claim 11, wherein the real value comprises a  
19 subband coefficient from a subband decomposition of a digitized signal.

20  
21 13. A method according to claim 4, wherein the step of decomposing the  
22 subband into outputs of quantizers comprises one of the following: a wavelet  
23 transform, a discrete cosine transform, and a motion compensation and  
24 discrete cosine based decomposition.

25  
26 14. A method according to claim 13, wherein the discrete cosine transform  
27 comprises a JPEG discrete cosine transform.

28

-51 -

1 15. A method according to claim 13, wherein the motion compensation  
2 and discrete cosine based decomposition comprise MPEG motion  
3 compensation and discrete cosine based decomposition.  
4

5 16. A method according to claim 1, wherein the digitized signal is one or  
6 more of a digital image, a digitized audio, a digitized video, and a digitized  
7 geometric representation of an object.  
8

9 17. A method according to claim 1, wherein the step of allocating forward  
10 error correction bits further comprises the step of modeling effects of network  
11 burst error to adjust the expected reconstructed image distortion.  
12

13 18. A method according to claim 17, wherein the step of modeling effects of  
14 network burst error comprises the step of applying a Markov model to predict  
15 expected packet losses between two or more transmitted packets.  
16

17 19. A method according to claim 17, further comprising the step of  
18 randomizing packet order in order to decorrelate burst losses during  
19 transmission.  
20

21 20. A method according to claim 1, wherein the step of allocating forward  
22 error correction bits further comprises the step of allocating bits between tasks  
23 of encoding image subbands and protecting encoded data with forward error  
24 correction.  
25

26 21. A method according to claim 20, wherein the step of allocating bits  
27 between tasks of encoding image subbands and protecting encoded data with  
28 forward error correction further comprises the step of determining a likely

1 distortion of the reconstructed image relative to compression and network  
2 losses, and subject to a total number of bytes transmitted.

3  
4 22. A method according to claim 1, wherein the step of allocating forward  
5 error correction bits further comprises the step of allocating a fixed number of  
6 bits between redundancy and data depending upon an expected loss rate  
7 through the Internet.

8  
9 23. A method according to claim 22, wherein the step of allocating a fixed  
10 number of bits further comprises the steps of shifting bits to redundancy for  
11 high loss rates, and shifting bits to data for lower loss rates.

12  
13 24. A method according to claim 1, further comprising the step of storing  
14 final transmission rates for one or more connections to reduce subsequent  
15 startup time to any of the connections.

16  
17 25. A method according to claim 1, further comprising the step of  
18 interleaving the subbands into smaller memory blocks.

19  
20 26. A method according to claim 25, wherein the step of interleaving the  
21 subbands into smaller memory blocks further comprises the step of forming  
22 memory blocks of up to about 150 bytes.

23  
24 27. A method according to claim 25, further comprising the step of  
25 compressing the memory blocks with an arithmetic decoder.

26  
27 28. A system for transmitting a digital image over the Internet, the digital  
28 image of the type that includes a series of subbands, comprising: means for  
29 allocating forward error correction bits and quantizer precision to each subband

-53 -

1 in order to reduce expected image distortion subject to an overall bit budget;  
2 and (b) means for transmitting the quantized image and forward error  
3 correction bits over the Internet.  
4

5 29. A system according to claim 28, further comprising means for  
6 decomposing each of the subbands into outputs of quantizers.  
7

8 30. A system according to claim 29, wherein the means for decomposing  
9 each of the subbands into outputs of quantizers comprises means for  
10 decomposing each of the subbands into outputs of nested quantizers.  
11

12 31. A system according to claim 29, wherein the means for decomposing  
13 each of the subbands into outputs of quantizers comprises one or more of the  
14 following: means for performing wavelet transform, means for performing a  
15 discrete cosine transform, and means for performing a motion compensation  
16 and discrete cosine based decomposition.  
17

18 32. A system for transmitting a digital image across the Internet,  
19 comprising: an image compression section for performing lossy image  
20 compression on the image; a bit allocation for source coding section for  
21 transforming the compressed image into a set of subbands, each subband being  
22 ranked relative to other subbands based upon its impact to image quality; a  
23 channel coding and expected image distortion section for allocating bits within  
24 the subbands between source and channel codes in order to minimize an  
25 expected reconstructed image distortion subject to an overall transmission bit  
26 budget; and means for transmitting a subset of the subbands with forward error  
27 correction bits on the Internet.  
28

-54 -

1 33. A system according to claim 32, wherein the image compression section  
2 comprises transform coder means for generating lossy compressed image.  
3

4 34. A system according to claim 16, wherein the transform coder means  
5 comprises one of JPEG, wavelet, wavelet packet, and DCT transforms.  
6

7 35. A system according to claim 32, wherein the image compression section  
8 comprises means for performing wavelet transforms.  
9

10 36. A system according to claim 35, wherein the means for performing  
11 wavelet transforms further comprises (a) means for quantizing the electronic  
12 image by quantizing the coefficients using uniform quantizers, and (b) an  
13 arithmetic coder for coding resulting coefficients for entropy.  
14

15 37. A system according to claim 32, wherein the bit allocation for source  
16 coding section comprises means for decomposing each of the subbands into  
17 outputs of quantizers.  
18

19 38. A system according to claim 37, wherein the means for decomposing  
20 each of the subbands into outputs of quantizers comprises one or more of the  
21 following: means for performing wavelet transform, means for performing a  
22 discrete cosine transform, and means for performing a motion compensation  
23 and discrete cosine based decomposition.  
24

25 39. A system according to claim 32, wherein the bit allocation for source  
26 coding section comprises means for dynamically allocating bits between source  
27 and channel codes depending upon conditions within a network.  
28



-55 -

1 40. A system according to claim 32, wherein the bit allocation for source  
2 coding section comprises means for finely quantizing a first group of  
3 coefficients, and coarsely quantizing a second group of coefficients, the first  
4 group having greater visual impact on image fidelity.

5  
6 41. A system according to claim 32, wherein the bit allocation for source  
7 coding section comprises means for determining a quantization resolution for  
8 each subband based upon a trade-off between quantization error, thereby  
9 allocating quantizer resolution to obtain a minimum image distortion for a  
10 given bit expenditure.

11  
12 42. A system according to claim 41, further comprising a rate control section  
13 having means for determining the bit expenditure based upon network  
14 conditions.

15  
16 43. A system according to claim 32, wherein the bit allocation for source  
17 coding section comprises means for decoding the image into a series of nested  
18 quantizers.

19  
20 44. A system according to claim 32, wherein the bit allocation for source  
21 coding section comprises means for determining which bitplanes to transmit  
22 and the redundancy applied to each bitplane.

23  
24 45. A system according to claim 32, wherein the bit allocation for source  
25 coding section comprises means for assessing image distortion.

26  
27 46. A system according to claim 45, wherein the means for assessing image  
28 distortion further comprises means for storing and assessing image distortion  
29 based upon a mean squared error function.

1  
2 47. A system according to claim 32, wherein the bit allocation for source  
3 coding section comprises means for refining quantizer resolution based upon  
4 marginal analysis.

5  
6 48. A system according to claim 32, wherein the channel coding and  
7 expected image distortion section comprises means for adaptively allocating  
8 quantizer resolution, thereby adding redundancy to transmitted images and  
9 optimizing a partition of bits between source and channel codes.

10  
11 49. A system according to claim 32, wherein the channel coding and  
12 expected image distortion section comprises means for interleaving subbands  
13 into smaller memory blocks.

14  
15 50. A system according to claim 49, wherein the means for interleaving  
16 subbands into smaller memory blocks comprises means for forming memory  
17 blocks of up to 150 bytes.

18  
19 51. A system according to claim 32, wherein the channel coding and  
20 expected image distortion section comprises means for compressing subbands  
21 with an arithmetic coder.

22  
23 52. A system according to claim 51, wherein the means for compressing  
24 subbands with an arithmetic coder comprises means for forming network  
25 packets having a fixed Internet size.

26  
27 53. A system according to claim 52, wherein the fixed Internet size is 576  
28 bytes.

1  
2 54. A system according to claim 51, wherein the channel coding and  
3 expected image distortion section comprises means for adding redundancy to  
4 compressed bitplanes by adding parity bits to a bitstream.

5  
6 55. A system according to claim 51, wherein the channel coding and  
7 expected image distortion section comprises means for interleaving bitplanes  
8 to reduce a visual impact of packet losses during transmission.

9  
10 56. A system according to claim 51, wherein the channel coding and  
11 expected image distortion section comprises means for determining a  
12 probability of network packet loss.

13  
14 57. A system according to claim 51, further comprising a rate control section  
15 for determining a round-trip travel time for transmission packets on the  
16 Internet.

17  
18 58. A method for transmitting an electronic image between a first node and  
19 a second node, each node being connected to the other through the Internet,  
20 each node comprising a digital data processor, comprising the steps of:  
21 compressing the electronic image with forward error correction at the first  
22 node so that image bits are concentrated within image portions having greater  
23 visual impact; transmitting the compressed electronic image on the Internet  
24 and between the first node and second node; and reconstructing the  
25 transmitted image at the second node by rebuilding fragments lost during  
26 transmission according to the forward error correction applied to the image.

27  
28 59. A method for transmitting an electronic image on the Internet,  
29 comprising the steps of: decomposing an electronically compressed image into

-58 -

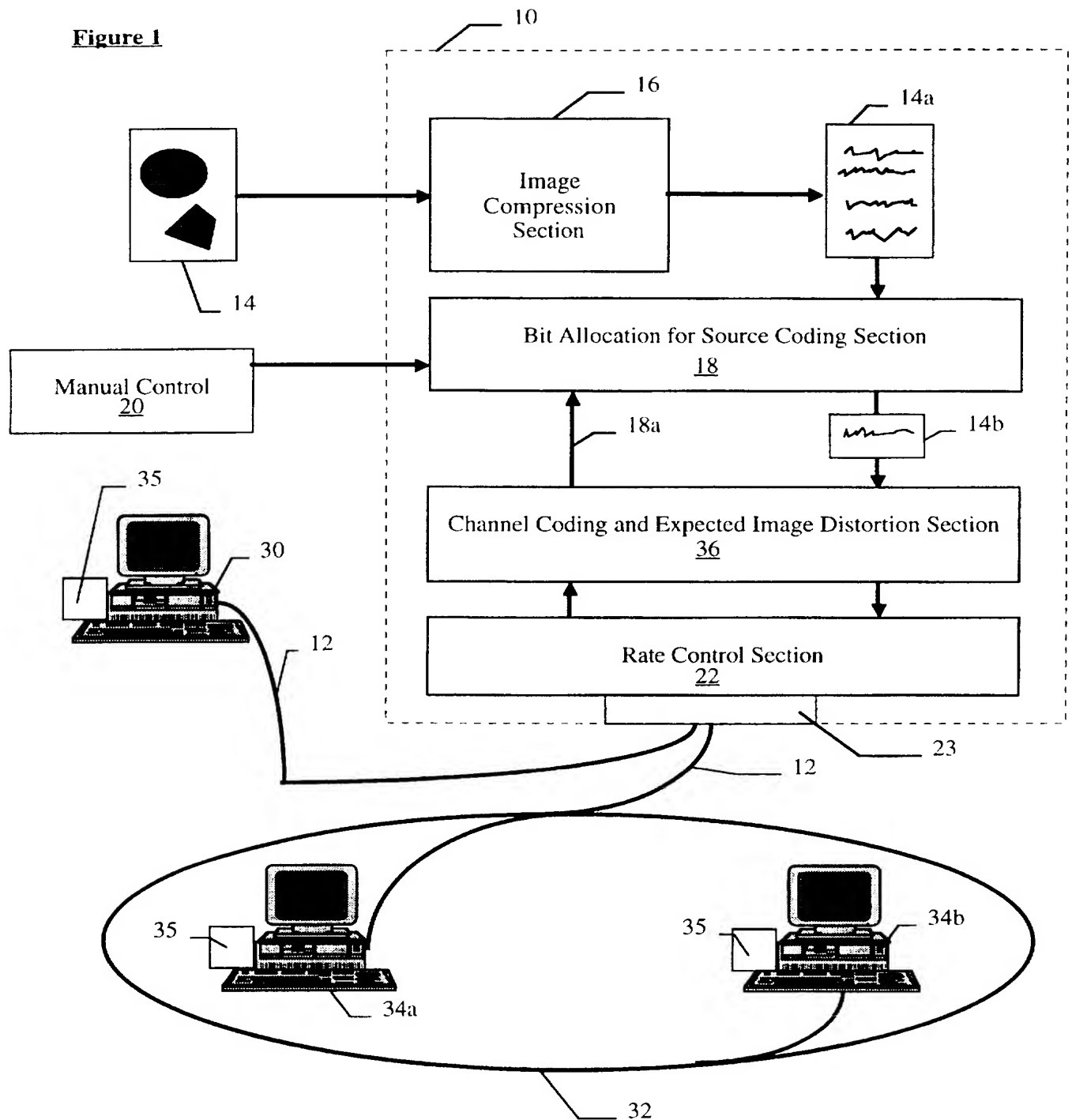
1 a series of subbands; selecting a subset of subbands to transmit on the Internet  
2 based upon a relative ranking between subbands; allocating forward error  
3 correction bits to each subband within the subset in order to minimize an  
4 expected reconstructed image distortion subject to an overall transmission bit  
5 budget; and transmitting the subset and the forward error correction bits  
6 within packets on the Internet.

7  
8 60. A method according to claim 59, further comprising the step of  
9 reconstructing the electronic image by (a) reading at least a portion of the  
10 packets from the Internet, and (b) decoding the subset of subbands according to  
11 the forward error correction bits allocated to the subbands, thereby rebuilding  
12 image fragments lost during transmission.

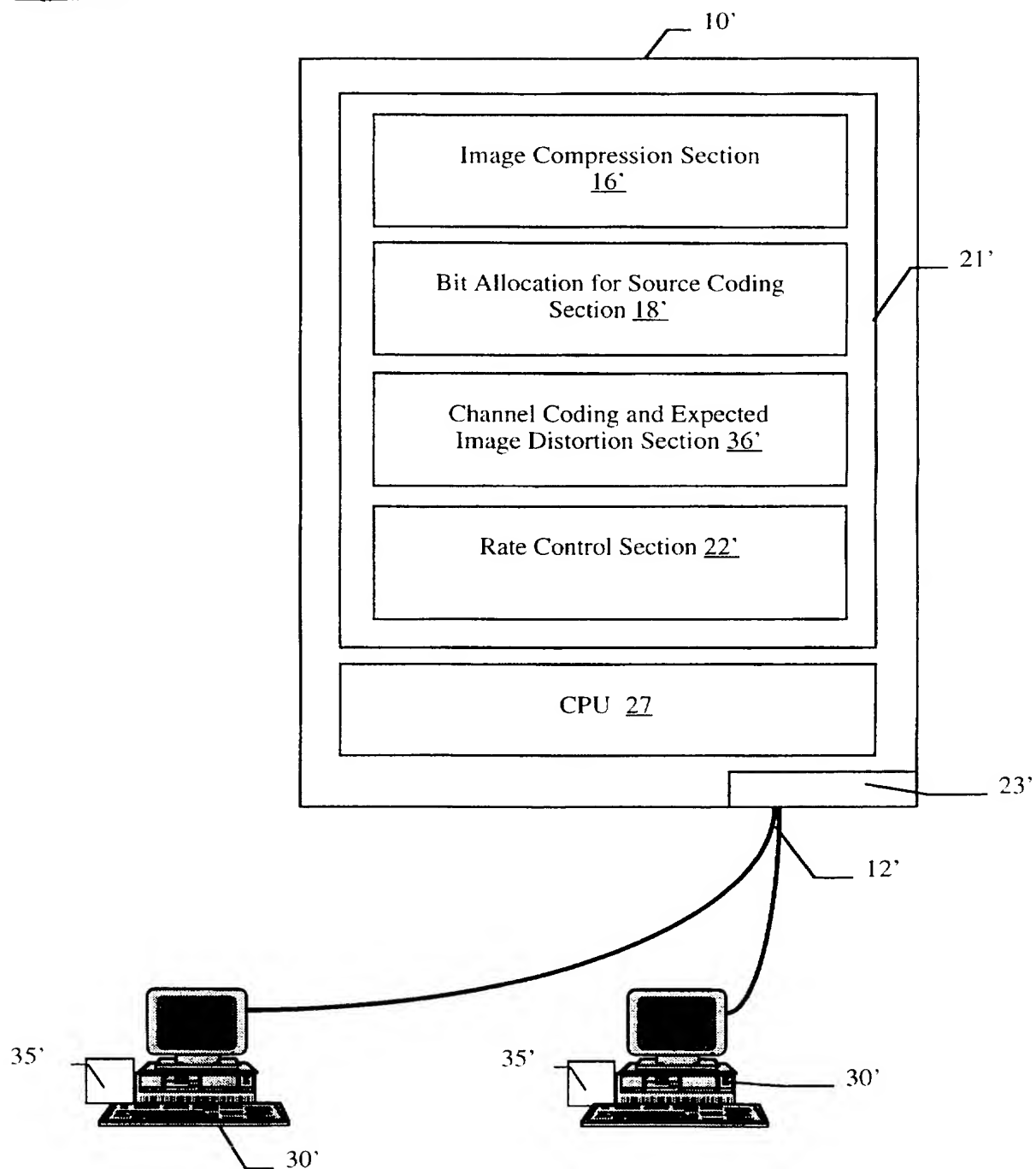
13  
14 61. A method according to claim 59, further comprising the steps of  
15 assessing a waiting period for transmitted packets and terminating the waiting  
16 period at an approximate time corresponding to an expected arrival time of a  
17 last packet plus a standard deviation of an interpacket arrival time.

18  
19  
20  
21

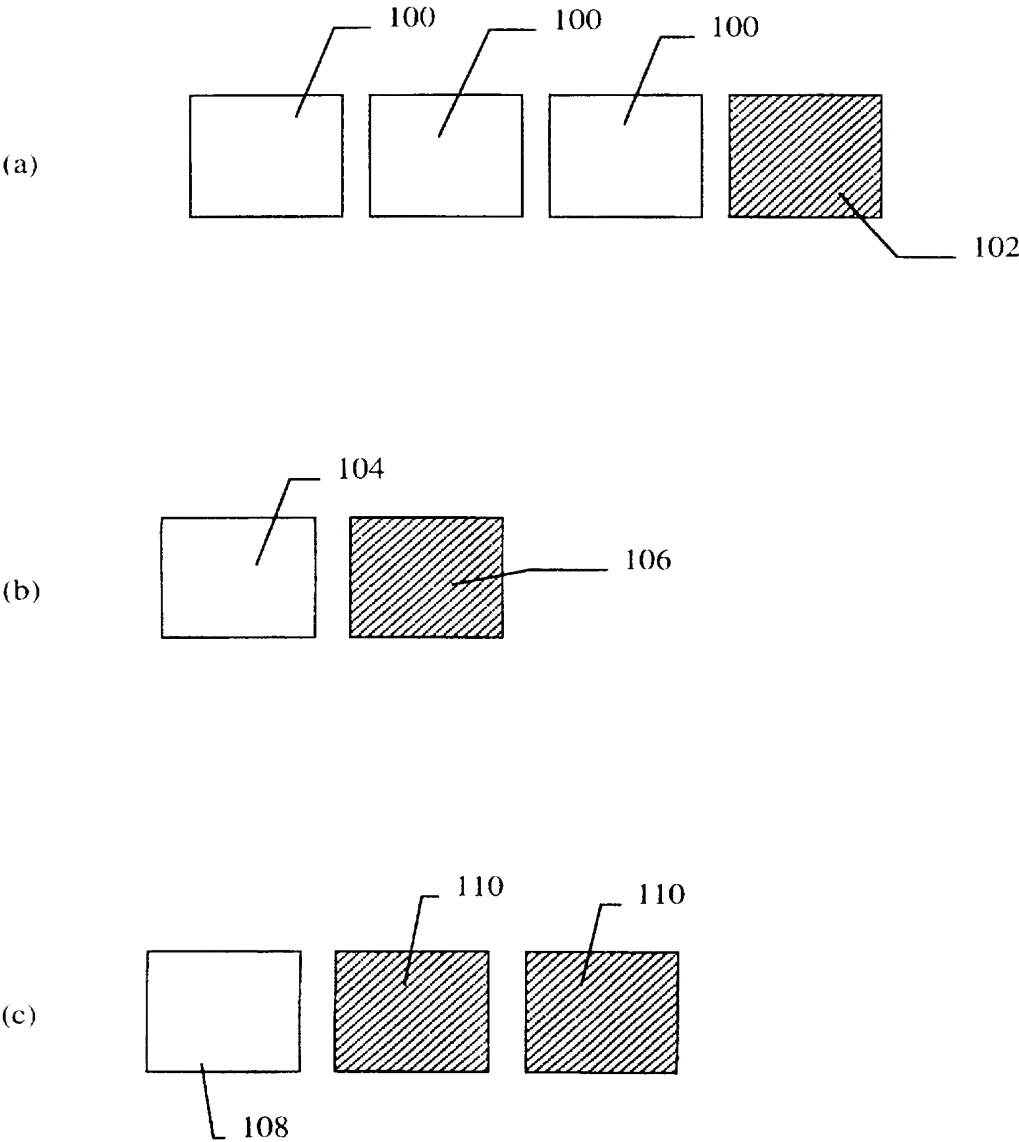
1/15

**Figure 1**

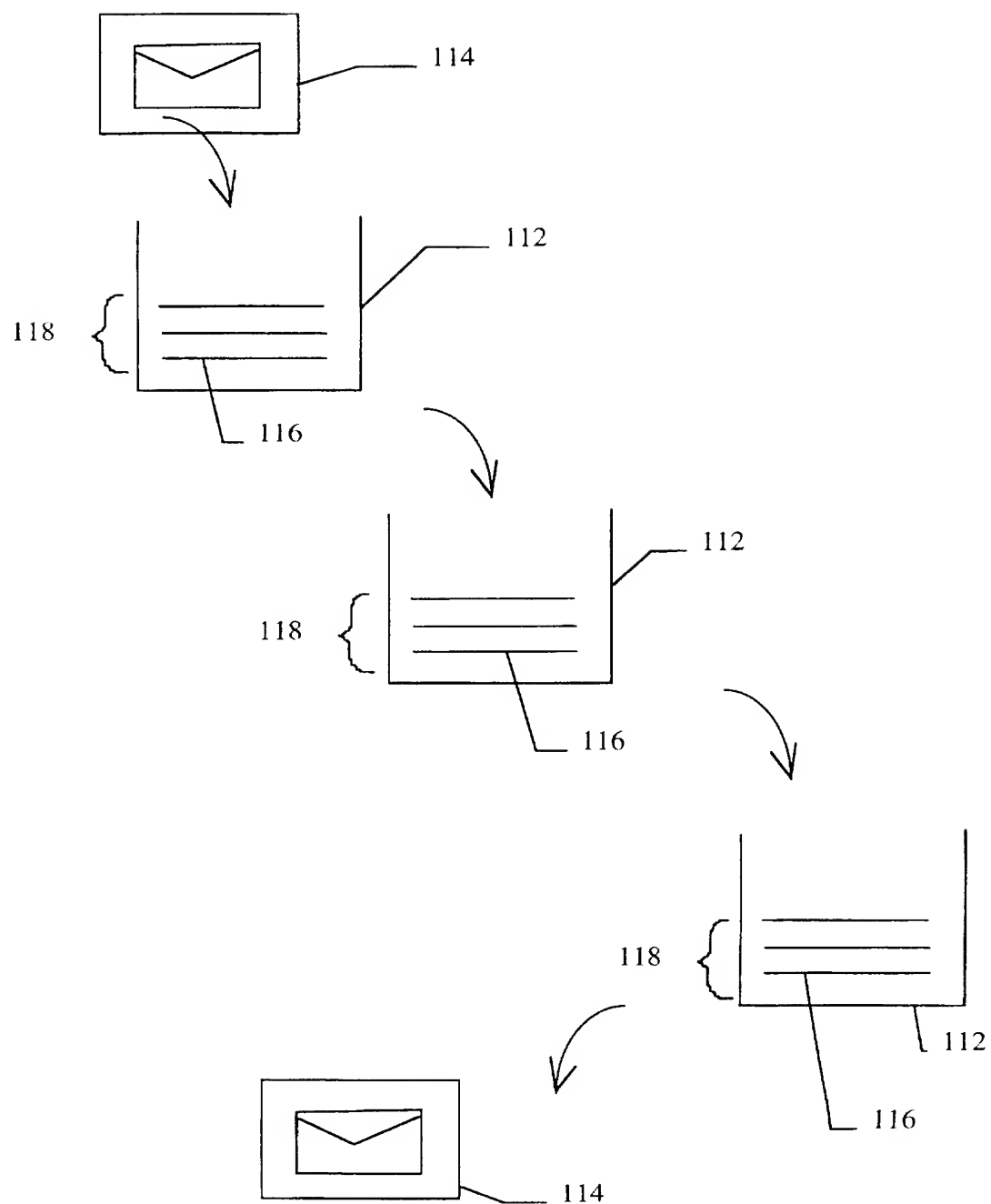
2/15

**Figure 1A**

**Figure 1B**



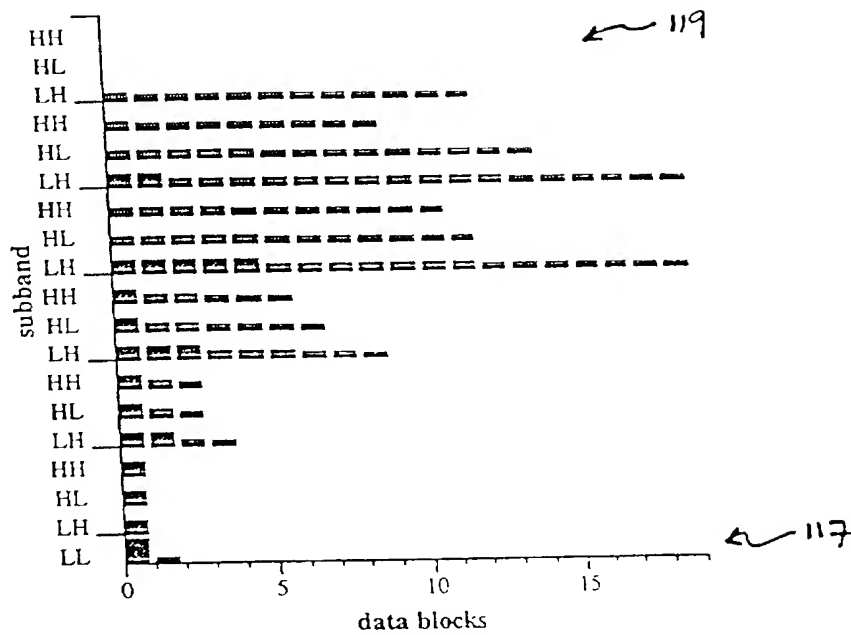
4/15

**Figure 1C**



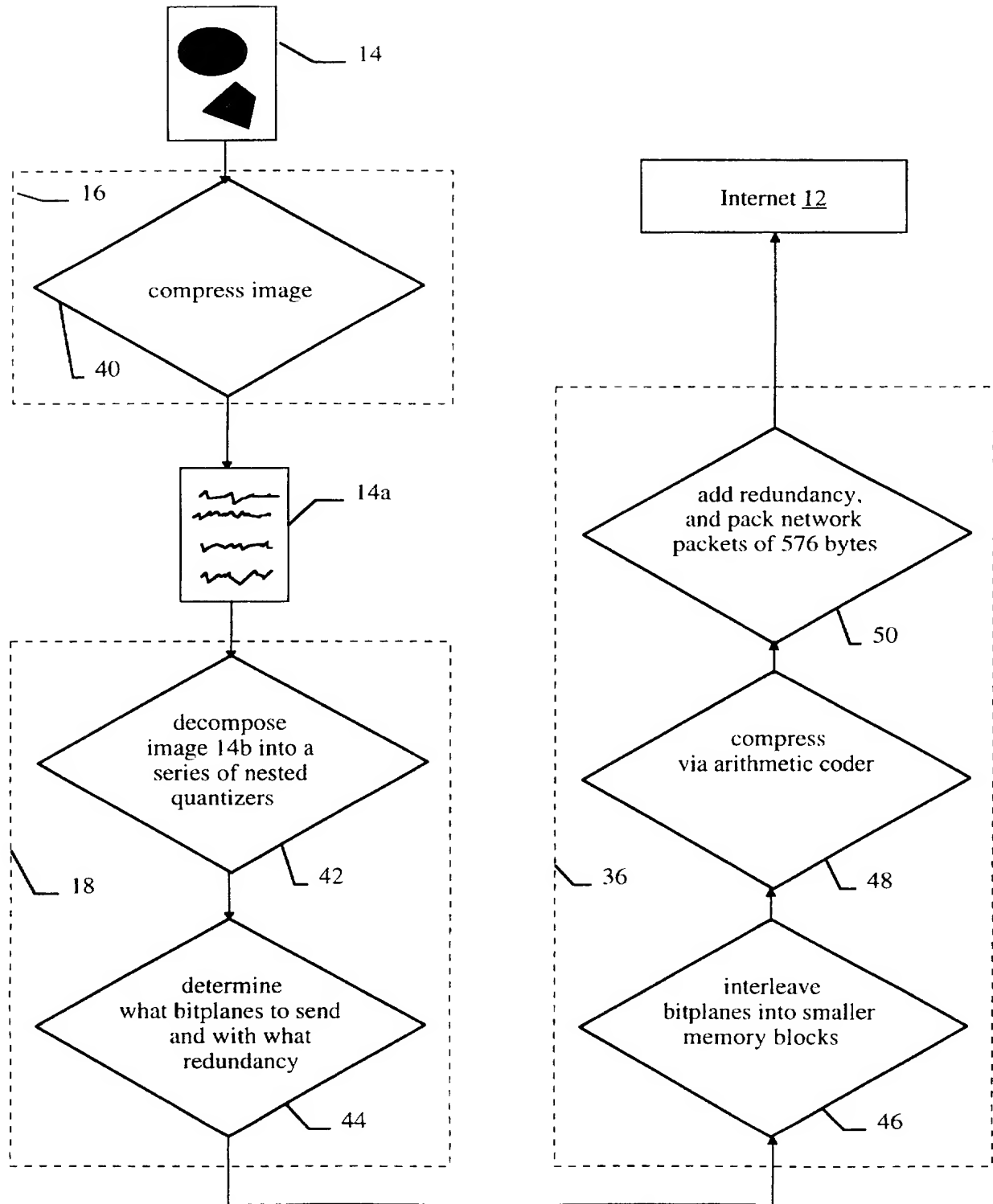
5/15

FIGURE 1D

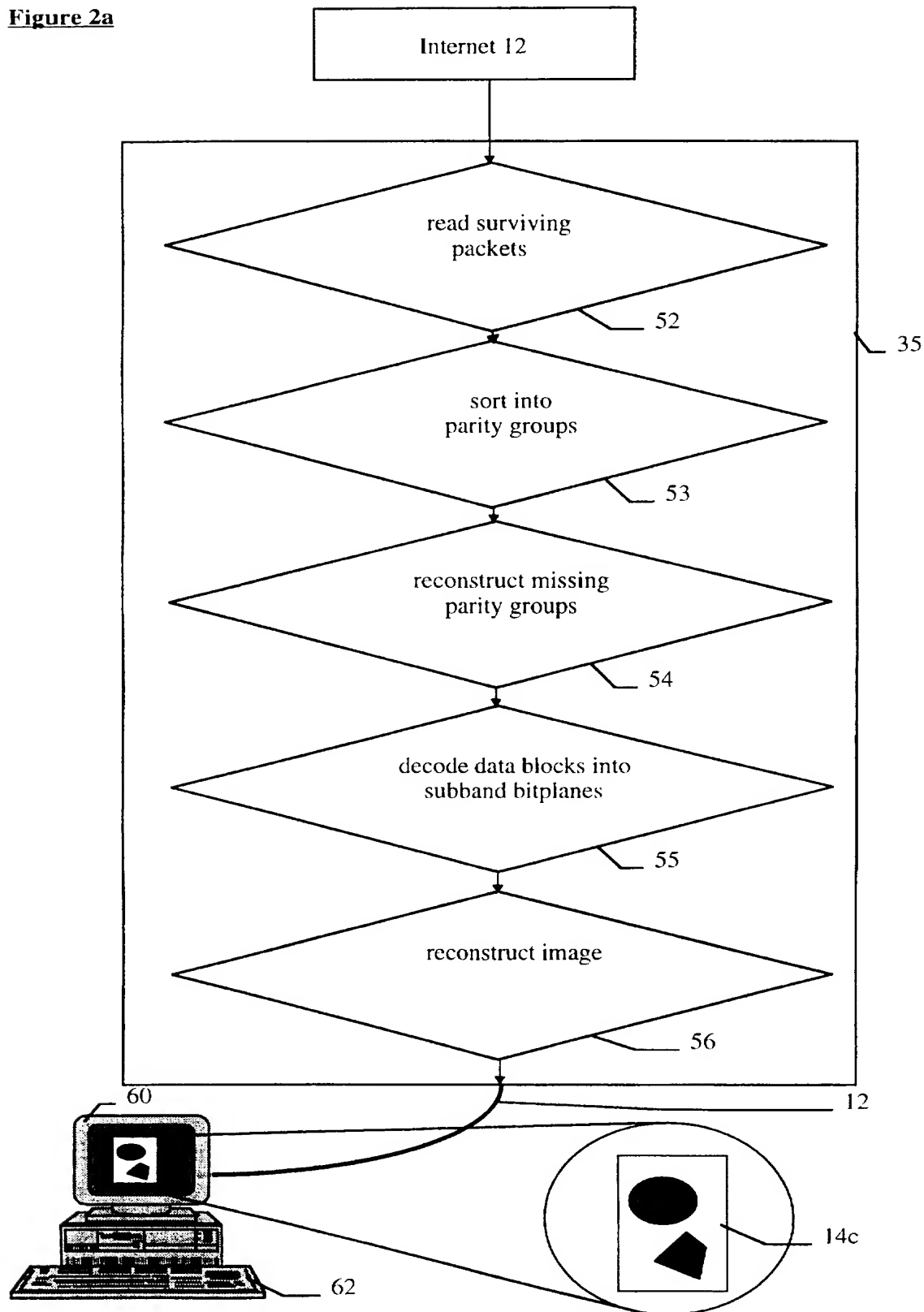


Uneven Allocation of Forward Error Correction Bits

6/15

**Figure 2**

7/15

**Figure 2a**

8/15

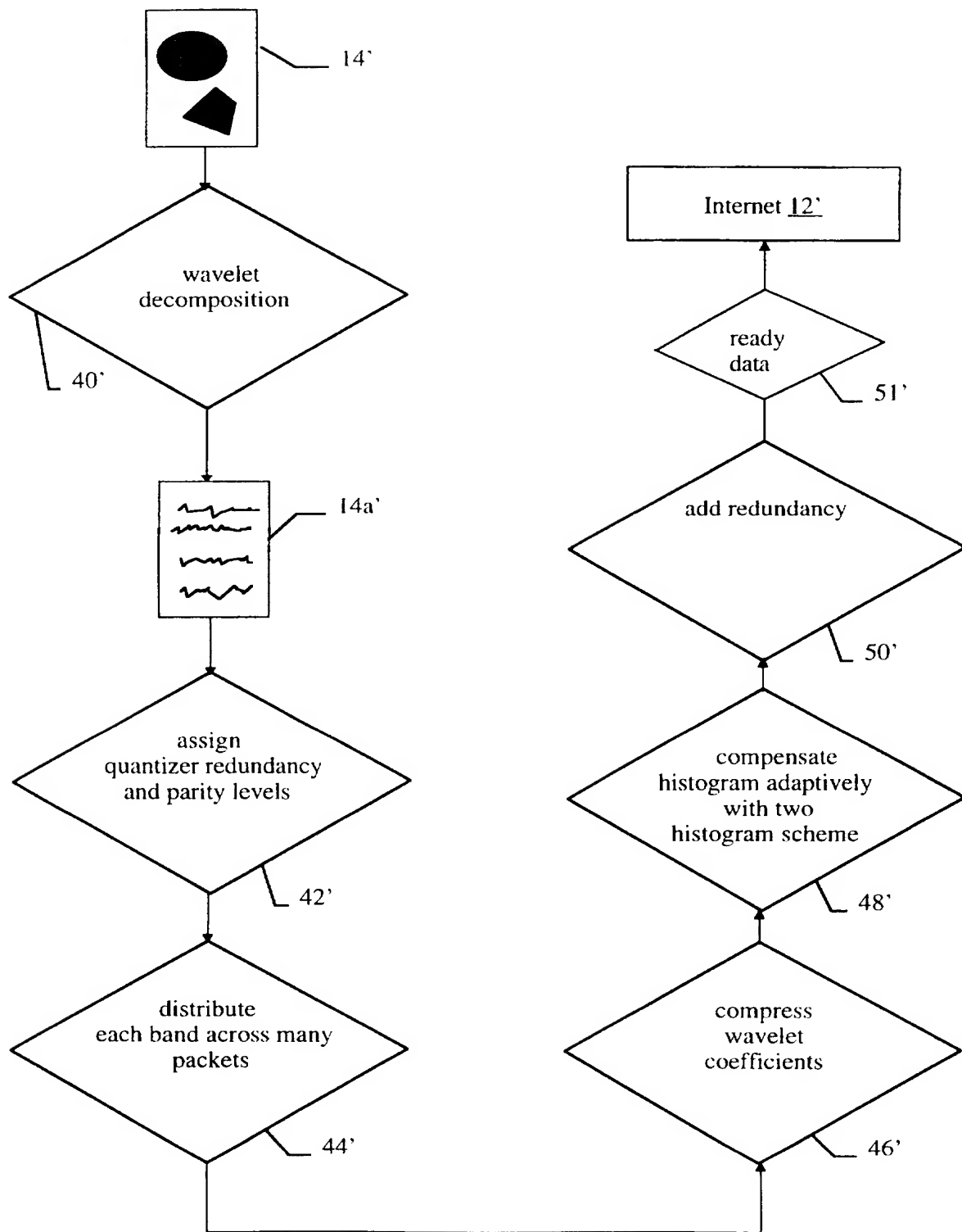
**Figure 2B**

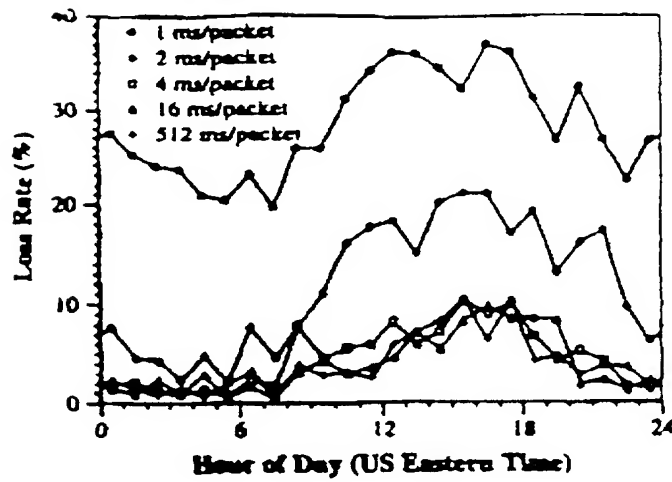
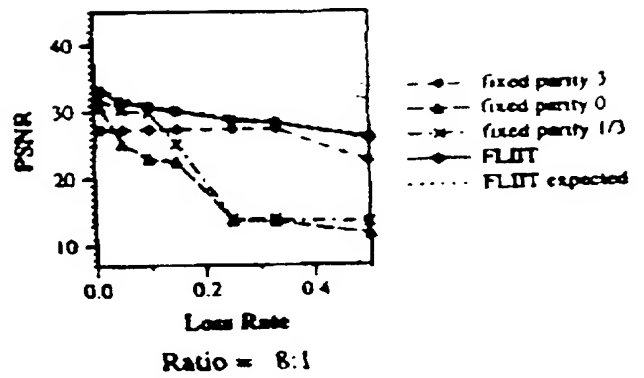
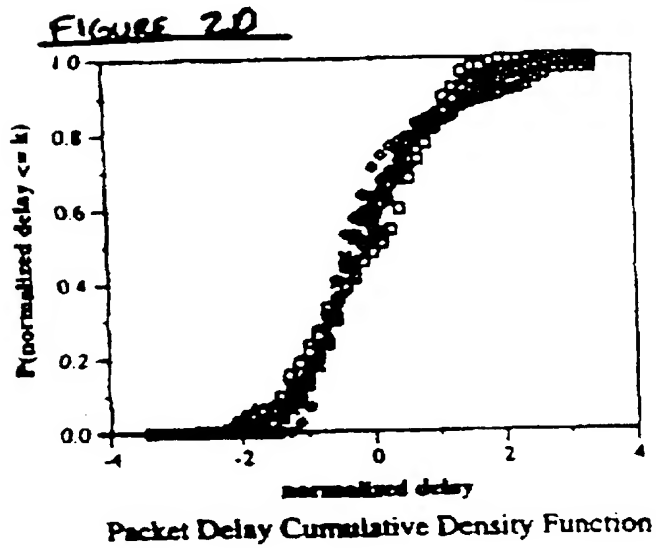
FIGURE 2CFIGURE 3



Fig. 4A



Fig. 5A



FIG. 4B



FIG. 5B

12/15



FIG. 4C



FIG. 5C





Fig. 5D

14/15

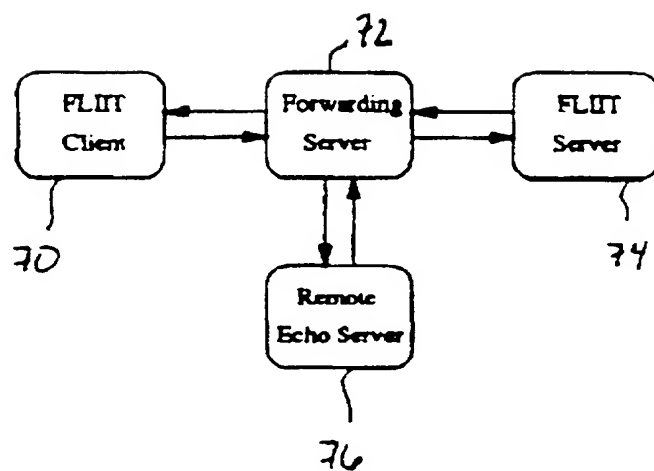
FIGURE 6

FIGURE 7

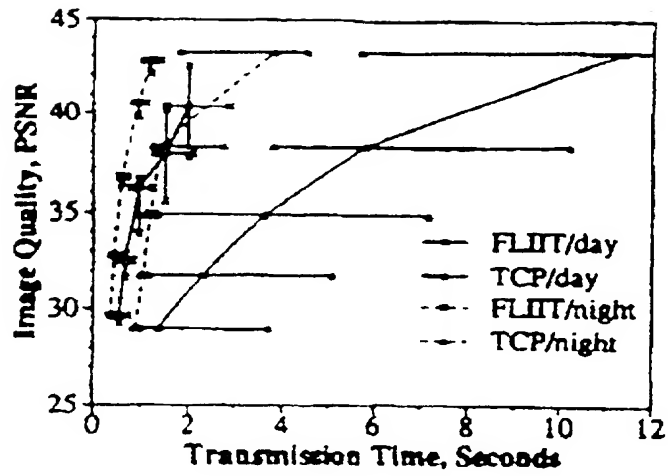


FIGURE 8A

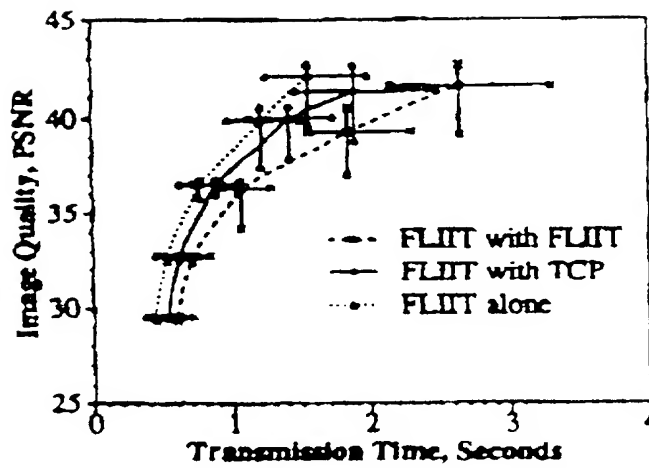
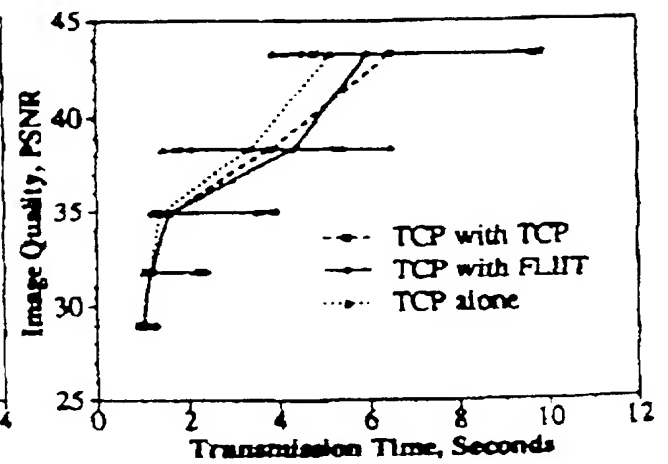


FIGURE 8B



## INTERNATIONAL SEARCH REPORT

 International application No.  
 PCT/US96/19388

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : HO4N 1/41

US CL : 358/426

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 358/426, 434; 348/12, 13

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,127,021 (SCHREIBER) 30 June 1992, Abstract, col. 9, line 58 - col. 10, line 8.	1-3, 16, 20, 22-23, 28, 32-34, 39, 58
Y	US 5,285,470 (SCHREIBER) 08 February 1994, Abstract, col. 11, lines 9-28.	1-3, 16, 20, 22-23, 28, 32-34, 39, 58
A	US 5,128,776 (SCORSE et al) 07 July 1992, Abstract, col. 8, lines 21-24.	1-61
A	US 5,208,682 (AHMED) 04 May 1993, Abstract, col. 3, lines 15-17.	1-61

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be part of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 22 APRIL 1997	Date of mailing of the international search report <b>13 MAY 1997</b>
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer <i>Stephen Brinich</i> STEPHEN BRINICH Telephone No. (703) 305 4390